Novel Proteins and Nucleic Acids Encoding Same

RELATED APPLICATIONS

This is a request for filing a new nonprovisional application under 37 C.F.R. §1.53(b). This application claims priority to U.S.S.N. 60/274,322 filed on March 8, 2001 (Cura 590); U.S.S.N. 60/283,675 filed on April 13, 2001 (Cura 590D1); U.S.S.N. 60/338,092 filed on December 3, 2001 (Cura 590D2); U.S.S.N. 60/274,281 filed on March 8, 2001 (Cura 591); U.S.S.N. 60/274,101 filed on March 8, 2001 (Cura 592); U.S.S.N. 60/325,681 filed on September 27, 2001 (Cura 592J1); U.S.S.N. 60/304,354 filed on July 10, 2001 (Cura 592I1); U.S.S.N. 60/279,995 filed on March 30, 2001 (Cura 592H1); U.S.S.N. 60/294,899 filed on May 31, 2001 (Cura 592E1); U.S.S.N. 60/287,424 filed on April 30, 2001 (Cura 592D1); U.S.S.N. 60/299,027 filed on June 18, 2001 (Cura 592D2); U.S.S.N. 60/309,198 filed on July 31, 2001 (Cura 592C1); U.S.S.N. 60/281,194 filed on April 4, 2001 (Cura 592A1); U.S.S.N. 60/274,194 filed on March 8, 2001 (Cura 593); U.S.S.N. 60/274,849 filed on March 9, 2001 (Cura 594); U.S.S.N. 60/330,380 filed on October 18, 2001 (Cura 594C1); U.S.S.N. 60/275,235 filed on March 12, 2001 (Cura 595); U.S.S.N. 60/288,342 filed on May 3, 2001 (Cura 595J1); U.S.S.N. 60/275,578 filed on March 13, 2001 (Cura 596); U.S.S.N. 60/291,240 filed on May 16, 2001 (Cura 596I1); U.S.S.N. 60/294,485 filed on May 30, 2001 (Cura 596B1); U.S.S.N. 60/299,310 filed on June 19, 2001 (Cura 596A1); U.S.S.N. 60/275,579 filed on March 13, 2001 (Cura 597); U.S.S.N. 60/275,601 filed on March 13, 2001 (Cura 598); U.S.S.N. 60/276,000 filed on March 14, 2001 (Cura 599); U.S.S.N. 60/280,900 filed on April 2, 2001 (Cura 599E1); U.S.S.N. 60/276,776 filed on March 16, 2001 (Cura 600); U.S.S.N. 60/294,889 filed on May 31, 2001 (Cura 600G1); U.S.S.N. 60/318,770 filed on September 12, 2001 (Cura 600E1); U.S.S.N. 60/276,994 filed on March 19, 2001 (Cura 604); U.S.S.N. 60/277,338 filed on March 20, 2001 (Cura 607); U.S.S.N. 60/325,430 filed on September 27, 2001 (Cura 607J1); U.S.S.N. 60/332,094 filed on November 21, 2001 (Cura 607C1); U.S.S.N. 60/299,303 filed on June 19, 2001 (Cura 607B1); U.S.S.N. 60/288,066 filed on May 2, 2001 (Cura 607A1); U.S.S.N. 60/277,321 filed on March 20, 2001 (Cura 608); U.S.S.N. 60/280,822 filed on April 2, 2001 (Cura 608A); U.S.S.N. 60/277,239 filed on March 20, 2001 (Cura 609); U.S.S.N. 60/277,327 filed on March 20, 2001 (Cura 610); U.S.S.N. 60/277,791 filed on March 21, 2001 (Cura 611); U.S.S.N. 60/333,184 filed on November 14, 2001 (Cura 611H1); U.S.S.N. 60/277,833 filed on March 22, 2001 (Cura 612); U.S.S.N. 60/318,462 filed on September 10, 2001 (Cura 612J1); U.S.S.N. 60/288,528 filed on May 3, 2001 (Cura 612A1); U.S.S.N.

1

60/278,152 filed on March 23, 2001 (Cura 613); U.S.S.N. 60/332,272 filed on November 14, 2001 (Cura 613D1); U.S.S.N. 60/278,894 filed on March 26, 2001 (Cura 614); U.S.S.N. 60/312,903 filed on August 16, 2001 (Cura 614C1); U.S.S.N. 60/333,272 filed on November 14, 2001 (Cura 614C2); U.S.S.N. 60/279,036 filed on March 27, 2001 (Cura 615); U.S.S.N. 60/332,172 filed on November 14, 2001 (Cura 615I1); U.S.S.N. 60/337,426 filed on December 3, 2001 (Cura 615I2); U.S.S.N. 60/278,999 filed on March 27, 2001 (Cura 616); U.S.S.N. 60/279,344 filed on March 28, 2001 (Cura 617); U.S.S.N. 60/332,271 filed on November 14, 2001 (Cura 617J1); U.S.S.N. 60/291,099 filed on May 16, 2001 (Cura 617H1); U.S.S.N. 60/291,190 filed on May 15, 2001 (Cura 617E1); U.S.S.N. 60/280,233 filed on March 30, 2001 (Cura 618); U.S.S.N. 60/280,802 filed on April 2, 2001 (Cura 621); U.S.S.N. 60/335,301 filed on October 31, 2001 (Cura 621F1); U.S.S.N. 60/337,185 filed on December 4, 2001 (Cura 621D1); and U.S.S.N. 60/345,705 filed on January 3, 2002 (Cura 621B1).

Field of the Invention

The present invention relates to novel polypeptides that are targets of small molecule drugs and that have properties related to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are gene products of novel genes, or are specified biologically active fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions. The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related proteins that are similar to them, are encoded by a cDNA and/or by genomic DNA. The proteins, polypeptides and their cognate nucleic acids were identified by Curagen Corporation in certain cases. The XYZase-encoded protein and any variants, thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for small molecule drugs. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify such therapeutic antibodies and/or therapeutic small molecules.

Background

Eukaryotic cells are characterized by biochemical and physiological processes which under normal conditions are exquisitely balanced to achieve the preservation and propagation of the cells. When such cells are components of multicellular organisms such as vertebrates, or more particularly organisms such as mammals, the regulation of the biochemical and physiological processes involves intricate signaling pathways. Frequently, such signaling pathways are constituted of extracellular signaling proteins, cellular receptors that bind the signaling proteins and signal transducing components located within the cells.

Signaling proteins may be classified as endocrine effectors, paracrine effectors or autocrine effectors. Endocrine effectors are signaling molecules secreted by a given organ into the circulatory system, which are then transported to a distant target organ or tissue. The target cells include the receptors for the endocrine effector, and when the endocrine effector binds, a signaling cascade is induced. Paracrine effectors involve secreting cells and receptor cells in close proximity to each other, for example two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the effector results in induction of the signaling cascade that elicits the corresponding biochemical or physiological effect. Autocrine effectors are highly analogous to paracrine effectors, except that the same cell type that secretes the autocrine effector also contains the receptor. Thus the autocrine effector binds to receptors on the same cell, or on identical neighboring cells. The binding process then elicits the characteristic biochemical or physiological effect.

Signaling processes may elicit a variety of effects on cells and tissues including by way of nonlimiting example induction of cell or tissue proliferation, suppression of growth or proliferation, induction of differentiation or maturation of a cell or tissue, and suppression of differentiation or maturation of a cell or tissue.

Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion protein effectors. In a clinical setting a subject may be suspected of suffering from a condition brought on by diminished or suppressed levels of a protein effector of interest. Therefore there is a need to be able to

3

assay for the level of the protein effector of interest in a biological sample from such a subject, and to compare the level with that characteristic of a nonpathological condition. There further is a need to provide the protein effector as a product of manufacture. Administration of the effector to a subject in need thereof is useful in treatment of the

5   pathological condition, or the protein effector deficiency or suppression may be favorably acted upon by the administration of another small molecule drug product. Accordingly, there is a need for a method of treatment of a pathological condition brought on by a diminished or suppressed levels of the protein effector of interest.

Small molecule targets have been implicated in various disease states or

10   pathologies. These targets may be proteins, and particularly enzymatic proteins, which are acted upon by small molecule drugs for the purpose of altering target function and achieving a desired result. Cellular, animal and clinical studies can be performed to elucidate the genetic contribution to the etiology and pathogenesis of conditions in which small molecule targets are implicated in a variety of physiologic, pharmacologic or native

15   states. These studies utilize the core technologies at CuraGen Corporation to look at differential gene expression, protein-protein interactions, large-scale sequencing of expressed genes and the association of genetic variations such as, but not limited to, single nucleotide polymorphisms (SNPs) or splice variants in and between biological samples from experimental and control groups. The goal of such studies is to identify potential

20   avenues for therapeutic intervention in order to prevent, treat the consequences or cure the conditions.

In order to treat diseases, pathologies and other abnormal states or conditions in which a mammalian organism has been diagnosed as being, or as being at risk for becoming, other than in a normal state or condition, it is important to identify new

25   therapeutic agents. Such a procedure includes at least the steps of identifying a target component within an affected tissue or organ, and identifying a candidate therapeutic agent that modulates the functional attributes of the target. The target component may be any biological macromolecule implicated in the disease or pathology. Commonly the target is a polypeptide or protein with specific functional attributes. Other classes of macromolecule

30   may be a nucleic acid, a polysaccharide, a lipid such as a complex lipid or a glycolipid; in addition a target may be a sub-cellular structure or extra-cellular structure that is comprised of more than one of these classes of macromolecule. Once such a target has been identified, it may be employed in a screening assay in order to identify favorable candidate therapeutic agents from among a large population of substances or compounds.

4

In many cases the objective of such screening assays is to identify small molecule candidates; this is commonly approached by the use of combinatorial methodologies to develop the population of substances to be tested. The implementation of high throughput screening methodologies is advantageous when working with large, combinatorial libraries

5   of compounds.

It is an objective of this invention to provide at least one target biopolymer that is intended to serve as the macromolecular component in a screening assay for identifying candidate pharmaceutical agents.

It is another objective of the present invention to provide screening assays that

10   positively identify candidate pharmaceutical agents from among a combinatorial library of low molecular weight substances or compounds.

It is still a further objective of this invention to employ the candidate pharmaceutical agents in any of a variety of in vitro, ex vivo and in vivo assays in order to identify pharmaceutical agents with advantageous therapeutic applications in the treatment

15   of a disease, pathology, or abnormal state or condition in a mammal.

Summary Of The Invention

The invention is based in part upon the discovery of nucleic acid sequences

20   encoding novel polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid, which represents the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, or polypeptide sequences, which represents the group consisting of SEQ ID NO: 2n, wherein n is an

25   integer between 1 and 178.

In one aspect, the invention provides an isolated polypeptide comprising a mature form of a NOVX amino acid. One example is a variant of a mature form of a NOVX amino acid sequence, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of

30   the mature form are so changed. The amino acid can be, for example, a NOVX amino acid sequence or a variant of a NOVX amino acid sequence, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also includes fragments of any of these. In another aspect, the invention also includes an

isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof.

Also included in the invention is a NOVX polypeptide that is a naturally occurring allelic variant of a NOVX sequence. In one embodiment, the allelic variant includes an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a NOVX nucleic acid sequence. In another embodiment, the NOVX polypeptide is a variant polypeptide described therein, wherein any amino acid specified in the chosen sequence is changed to provide a conservative substitution. In one embodiment, the invention discloses a method for determining the presence or amount of the NOVX polypeptide in a sample. The method involves the steps of: providing a sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the NOVX polypeptide, thereby determining the presence or amount of the NOVX polypeptide in the sample. In another embodiment, the invention provides a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide in a mammalian subject. This method involves the steps of: measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in the sample of the first step to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

In a further embodiment, the invention includes a method of identifying an agent that binds to a NOVX polypeptide. This method involves the steps of: introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. In various embodiments, the agent is a cellular receptor or a downstream effector.

In another aspect, the invention provides a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a NOVX polypeptide. The method involves the steps of: providing a cell expressing the NOVX polypeptide and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a

6

composition devoid of the substance, the substance is identified as a potential therapeutic agent. In another aspect, the invention describes a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with the NOVX polypeptide. This method involves the following steps: administering a test compound to a

5   test animal at increased risk for a pathology associated with the NOVX polypeptide, wherein the test animal recombinantly expresses the NOVX polypeptide. This method involves the steps of measuring the activity of the NOVX polypeptide in the test animal after administering the compound of step; and comparing the activity of the protein in the test animal with the activity of the NOVX polypeptide in a control animal not administered

10   the polypeptide, wherein a change in the activity of the NOVX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the NOVX polypeptide. In one embodiment, the test animal is a recombinant test animal that expresses a test protein transgene or expresses the transgene under the control of a promoter at an increased level

15   relative to a wild-type test animal, and wherein the promoter is not the native gene promoter of the transgene. In another aspect, the invention includes a method for modulating the activity of the NOVX polypeptide, the method comprising introducing a cell sample expressing the NOVX polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

20   The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. In a preferred embodiment, the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant. In another embodiment, the nucleic acid encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally

25   occurring polypeptide variant. In another embodiment, the nucleic acid molecule differs by a single nucleotide from a NOVX nucleic acid sequence. In one embodiment, the NOVX nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, or a complement of the nucleotide sequence. In another aspect, the invention

30   provides a vector or a cell expressing a NOVX nucleotide sequence.

In one embodiment, the invention discloses a method for modulating the activity of a NOVX polypeptide. The method includes the steps of: introducing a cell sample expressing the NOVX polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide. In another embodiment, the

7

invention includes an isolated NOVX nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising a NOVX amino acid sequence or a variant of a mature form of the NOVX amino acid sequence, wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes an amino acid sequence that is a variant of the NOVX amino acid sequence, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed.

In one embodiment, the invention discloses a NOVX nucleic acid fragment encoding at least a portion of a NOVX polypeptide or any variant of the polypeptide, wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed. In another embodiment, the invention includes the complement of any of the NOVX nucleic acid molecules or a naturally occurring allelic nucleic acid variant. In another embodiment, the invention discloses a NOVX nucleic acid molecule that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant. In another embodiment, the invention discloses a NOVX nucleic acid, wherein the nucleic acid molecule differs by a single nucleotide from a NOVX nucleic acid sequence.

In another aspect, the invention includes a NOVX nucleic acid, wherein one or more nucleotides in the NOVX nucleotide sequence is changed to a different nucleotide provided that no more than 15% of the nucleotides are so changed. In one embodiment, the invention discloses a nucleic acid fragment of the NOVX nucleotide sequence and a nucleic acid fragment wherein one or more nucleotides in the NOVX nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed. In another embodiment, the invention includes a nucleic acid molecule wherein the nucleic acid molecule hybridizes under stringent conditions to a NOVX nucleotide sequence or a complement of the NOVX nucleotide sequence. In one embodiment, the invention includes a nucleic acid molecule, wherein the sequence is changed such that no more than 15% of the nucleotides in the coding sequence differ from the NOVX nucleotide sequence or a fragment thereof.

8

In a further aspect, the invention includes a method for determining the presence or amount of the NOVX nucleic acid in a sample. The method involves the steps of: providing the sample; introducing the sample to a probe that binds to the nucleic acid molecule; and determining the presence or amount of the probe bound to the NOVX

5   nucleic acid molecule, thereby determining the presence or amount of the NOVX nucleic acid molecule in the sample. In one embodiment, the presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

In another aspect, the invention discloses a method for determining the presence of or predisposition to a disease associated with altered levels of the NOVX nucleic acid

10   molecule of in a first mammalian subject. The method involves the steps of: measuring the amount of NOVX nucleic acid in a sample from the first mammalian subject; and comparing the amount of the nucleic acid in the sample of step (a) to the amount of NOVX nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease; wherein an alteration in the level of the nucleic

15   acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those

20   described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be

25   limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

30   **Detailed Description Of The Invention**

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded

35   polypeptides, antibodies, and other related compounds. The sequences are collectively

9

referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table 1 provides a summary of the NOVX nucleic acids and

5   their encoded polypeptides.

TABLE 1. Sequences and Corresponding SEQ ID Numbers

NOVX No.	Internal Acc. No.	Nucleic Acid SEQ ID NO.	Amino Acid SEQ ID NO.	Homology
1a	CG58522-01	1	2	human platelet activating factor acetylhydrolase
2a	CG58520-01	3	4	GABA(A) receptor
2b	CG58520-02	5	6	GABA(A) receptor
2c	CG58520-03	7	8	GABA(A) receptor
3a	CG58518-01	9	10	GABA(A) receptor
4a	CG58516-01	11	12	Beta transducin
5a	CG58473-01	13	14	Protein kinase
6a	CG58470-01	15	16	UDP-N-acetylhexosamine pyrophosphorylase
7a	CG58593-01	17	18	ubiquitin 52 like
8a	CG57871-01	19	20	tousled like kinase like
9a	CG58590-01	21	22	guanylate kinase like
9b	CG58590-02	23	24	guanylate kinase like
10a	CG58572-01	25	26	glucosamine phosphate N acetyltransferase like
10b	CG58572-02	27	28	glucosamine phosphate N acetyltransferase like
11a	CG58564-01	29	30	Protein tyrosine phosphatase like
11b	CG58564-02	31	32	Protein tyrosine phosphatase like
11c	CG58564-03	33	34	Dual-Specificity phosphatase like
11d	CG58564-04	35	36	Dual-Specificity phosphatase like
12a	CG57819-01	37	38	RPGR interacting protein 1 like
13a	CG57789-01	39	40	RAS like protein RRP22 like
13b	CG57789-02	41	42	RAS like protein RRP22 like
14a	CG57758-01	43	44	sodium/lithium dependent dicarboxylate transporter like
14b	CG57758-02	45	46	sodium/lithium

				dependent dicarboxylate transporter like
14c	CG57758-03	47	48	sodium/lithium dependent dicarboxylate transporter like
14d	CG57758-04	49	50	sodium/lithium dependent dicarboxylate transporter like
14e	CG57758-05	51	52	sodium/lithium dependent dicarboxylate transporter like
15a	CG57732-01	53	54	Ca 2+ calmodulin dependent protein kinase IV kinase like
15b	CG57732-02	55	56	Ca 2+ calmodulin dependent protein kinase IV kinase like
15c	CG57732-03	57	58	Ca 2+ calmodulin dependent protein kinase IV kinase like
16a	CG57709-01	59	60	TCE2 like
17a	CG57700-01	61	62	hydoxyacylglutathione hydrolase like
17b	CG57700-02	63	64	hydoxyacylglutathione hydrolase like
17c	CG57700-03	65	66	hydoxyacylglutathione hydrolase like
17d	CG57700-04	67	68	hydoxyacylglutathione hydrolase like
18a	CG58553-01	69	70	vasopressin receptor like
19a	CG58626-01	71	72	phosphatidic acid preferring phospholipase A1 like
20a	CG57597-01	73	74	hypothetical protein like
21a	CG57804-01	75	76	Talin like
22a	CG57551-01	77	78	NAC-1 like
23a	CG57411-01	79	80	Kelch like
24a	CG57399-01	81	82	phospholipase ADRAB-B precursor like
24b	CG57399-02	83	84	phospholipase ADRAB-B precursor like
24c	CG57399-03	85	86	phospholipase ADRAB-B precursor like
25a	CG59311-01	87	88	acyl-coenzyme A thioester hydrolase

25b	CG59311-02	89	90	peroxisomal acyl-coenzyme A thioeseter hydrolase like
25c	CG59311-03	91	92	peroxisomal acyl-coenzyme A thioeseter hydrolase like
26a	CG59309-01	93	94	acyl-coenzyme A thioester hydrolase
27a	CG57364-01	95	96	CG6896
28a	CG59348-01	97	98	cytoplasmic protein (patent calls this Cyclin L-like)
29a	CG59245-01	99	100	glucose 6-phosphatase
29b	CG59245-02	101	102	glucose 6-phosphatase
30a	CG59241-01	103	104	Amiloride-sensitive sodium channel
31a	CG58602-01	105	106	FAD binding domain containing protein
32a	CG58468-01	107	108	Serum Amyloid Protein
33a	CG58183-01	109	110	N-Methyl-D-Aspartate receptor
34a	CG59315-01	111	112	Connexin
35a	CG59203-01	113	114	lysozyme C
35b	CG59203-02	115	116	lysozyme C
36a	CG58662-01	117	118	cytoplasmic protein
36b	CG58662-02	119	120	cytoplasmic protein
37a	CG58584-01	121	122	40S ribosomal protein S29 like
38a	CG58538-01	123	124	Histone deacetylase complex protein 66 like
39a	CG59371-01	125	126	expressed cytoplasmic protein like
40a	CG59346-01	127	128	cortactin binding protein 1 like
41a	CG57814-01	129	130	Basic I 19 like homo sapiens
41b	CG57814-02	131	132	Basic I 19 like homo sapiens
42a	CG59327-01	133	134	Monocarboxylate transporter 1 like
43a	CG59494-01	135	136	myelin P2 like
44a	CG59432-01	137	138	chloride channel like
44b	CG59432-02	139	140	chloride channel like
45a	CG59394-01	141	142	GPCR like
46a	CG59383-01	143	144	D6MM5E PROTEIN like
46b	CG59383-02	145	146	D6MM5E PROTEIN like
47a	CG58526-01	147	148	scramblase like
48a	CG57851-01	149	150	sulfotransferase like
49a	CG59377-01	151	152	epsin like
50a	CG59258-01	153	154	transcriptional activator like
51a	CG59492-01	155	156	Myosin Head (Motor

				Domain) like
52a	CG59564-01	157	158	Sorting nexin 6 like
53a	CG59553-01	159	160	Secretory protein SEC8 like
54a	CG59545-01	161	162	Placental protein 13 like
55a	CG59435-01	163	164	Nedd-1 like
55b	CG59435-02	165	166	Nedd-1 like
56a	CG59439-01	167	168	Xenobiotic/medium-chain fatty acid:CoA ligase form XL-III like
56b	CG59439-02	169	170	Xenobiotic/medium-chain fatty acid:CoA ligase form XL-III like
57a	CG59354-01	171	172	phosducin like
57b	CG59354-02	173	174	phosducin like
57c	CG59354-03	175	176	phosducin like
58a	CG59319-01	177	178	phosducin like
58b	CG59319-02	179	180	phosducin like
59a	CG59576-01	181	182	GPCR like
60a	CG59557-01	183	184	GPCR like
61a	CG59555-01	185	186	GPCR like
62a	CG59551-01	187	188	GPCR like
63a	CG59540-01	189	190	GPCR like
64a	CG59280-01	191	192	GPCR like
64b	CG59280-02	193	194	GPCR like
65a	CG59568-01	195	196	GPCR like
66a	CG59224-01	197	198	GPCR like
67a	CG59222-01	199	200	GPCR like
68a	CG59220-01	201	202	GPCR like
69a	CG59218-01	203	204	GPCR like
70a	CG59216-01	205	206	GPCR like
71a	CG59214-01	207	208	GPCR like
72a	CG59211-01	209	210	GPCR like
73a	CG59276-01	211	212	Dihydroorotate dehydrogenase like
74a	CG59268-01	213	214	monooxygenase like
75a	CG59549-01	215	216	H326 like (cytoplasmic protein with WD repeat domain)
76a	CG59641-01	217	218	Acetyl-CoA Carboxylase 2 like
77a	CG59630-01	219	220	Midnolin like
78a	CG59561-01	221	222	ACYL COENZYME A THIOESTER HYDROLASE like
79a	CG59452-01	223	224	CELL PROLIFERATION RELATED PROTEIN CAP like
80a	CG59572-01	225	226	Pseudouridine Synthase 3 like
80b	CG59572-02	227	228	Pseudouridine Synthase 3 like

13

81a	CG59522-01	229	230	Myosin like
82a	CG59520-01	231	232	Farnesyl-pyrophosphate synthetase like
83a	CG59758-01	233	234	UBIQUITIN like
83b	CG59758-02	235	236	UBIQUITIN like
84a	CG59586-01	237	238	glucokinase like
85a	CG59704-01	239	240	serine/threonine kinase like
86a	CG59628-01	241	242	Short-chain dehydrogenase like
87a	CG59516-01	243	244	Calponin like
87b	CG59516-02	245	246	Calponin like
88a	CG59671-02	247	248	acyl-coenzyme A thioester hydrolase
89a	CG56870-01	249	250	NDRG3 like
89b	CG56870-02	251	252	NDRG3 like
89c	CG56870-03	253	254	NDRG3 like
89d	CG56870-04	255	256	NDRG3 like
89e	CG56870-05	257	258	NDRG3 like
90a	CG59764-01	259	260	Ferritin like
91a	CG59710-01	261	262	P14 like
92a	CG59754-02	263	264	Downs syndrome cell adhesion molecule like
92b	CG59754-01	265	266	Downs syndrome cell adhesion molecule like
93a	CG59800-01	267	268	HEPARAN SULFATE D-GLUCOSAMINYL 3-O-SULFOTRANSFERASE-3B like
94a	CG59761-01	269	270	AXIN 1 (AXIS INHIBITION PROTEIN 1) (HAXIN) like
95a	CG59756-01	271	272	JUNCTOPHILIN TYPE 2 like
96a	CG59708-01	273	274	Ubiquitin carboxyl-terminal hydrolase 21 like
96b	CG59708-02	275	276	Ubiquitin carboxyl-terminal hydrolase 21 like
96c	CG59708-03	277	278	Ubiquitin carboxyl-terminal hydrolase 21 like
97a	CG59559-01	279	280	BA12M19.1.3 like
98a	CG59669-01	281	282	carbonyl reductase (called NADPH-dependent carbonyl reductase-like in patent)
99a	CG58624-01	283	284	metal transporter
100a	CG59679-01	285	286	carbonyl reductase

14

101a	CG59644-01	287	288	CG12091 (putative protein phosphatase)
102a	CG59662-01	289	290	Cyclophilin
103a	CG59773-01	291	292	Myomegalin
103b	CG59773-02	293	294	Myomegalin
103c	CG59773-03	295	296	Myomegalin
104a	CG57460-01	297	298	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE like
105a	CG57464-01	299	300	N-ACETYLTRANSFER ASE like
106a	CG57466-01	301	302	Acetylglucosaminyltra nsferase like
107a	CG57468-01	303	304	ABC transporter like homo sapiens
108a	CG59609-01	305	306	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A like
109a	CG59613-01	307	308	Proliferating cell nuclear antigen like
110a	CG59619-01	309	310	CYTOPLASMIC ACTIN 2 like
111a	CG59621-01	311	312	SELENOPHOSPHAT E SYNTHETASE like
112a	CG59625-01	313	314	glucose transporter like
113a	CG59887-01	315	316	Amino Acid/Metabolite Permease like
113b	CG59887-02	317	318	Amino Acid/Metabolite Permease like
114a	CG59861-01	319	320	RIBULOSE-5-PHOSPHATE-EPIMERASE like
114b	CG59861-02	321	322	RIBULOSE-5-PHOSPHATE-EPIMERASE like
115a	CG59857-01	323	324	Rhotekin like homo sapiens
116a	CG59855-01	325	326	ATP SYNTHASE SUBUNIT C lik
116b	CG59855-02	327	328	ATP SYNTHASE SUBUNIT C like
117a	CG59807-01	329	330	Zinc finger like
118a	CG59805-01	331	332	Zinc finger like
119a	CG59928-01	333	334	Universal Stress (USP) Domain Containing Protein like
120a	CG59947-01	335	336	VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN KV3.3 (KSHIIID) like
121a	CG59938-01	337	338	arylsulfatase like homo sapiens
122a	CG59746-01	339	340	ubiquitin-specific

15

				processing protease like homo sapiens
123a	CG88613-01	341	342	INOSITOL 1,4,5-TRISPHOSPHATE 3-KINASE ISOENZYME like
124a	CG59993-01	343	344	synaptotagmin II like
124b	CG59993-02	345	346	synaptotagmin II like
125a	CG59991-01	347	348	ooplasm specific protein like
126a	CG59987-01	349	350	GTP-RHO binding protein 1 (rhophilin) like
126b	CG59987-02	351	352	GTP-RHO binding protein 1 (rhophilin) like
127a	CG59971-01	353	354	Leucine rich repeat (LRR) like
127b	CG59971-02	355	356	Leucine rich repeat (LRR) like

Table 1 indicates homology of NOVX nucleic acids to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table 1 will be useful in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table 1.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table 1, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the

16

identification of small molecules that modulate or inhibit diseases associated with the protein families listed in Table 1.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C.

5   Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, e.g.a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the
10   invention are disclosed herein.

The present invention is based on the identification of biological macromolecules differentially modulated in a pathologic state, disease, or an abnormal condition or state. Among the pathologies or diseases of present interest include metabolic diseases including those related to endocrinologic disorders, cancers, various tumors and neoplasias,
15   inflammatory disorders, central nervous system disorders, and similar abnormal conditions or states. In very significant embodiments of the present invention, the biological macromolecules implicated in the pathologies and conditions are proteins and polypeptides, and in such cases the present invention is related as well to the nucleic acids that encode them. Methods that may be employed to identify relevant biological
20   macromolecules include any procedures that detect differential expression of nucleic acids encoding proteins and polypeptides associated with the disorder, as well as procedures that detect the respective proteins and polypeptides themselves. Significant methods that have been employed by the present inventors, include GeneCalling ® technology and SeqCalling TM technology, disclosed respectively, in U. S. Patent No. 5,871,697, and in U.
25   S. Ser. No. 09/417,386, filed Oct. 13, 1999, each of which is incorporated herein by reference in its entirety. GeneCalling ® is also described in Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999).

The invention provides polypeptides and nucleotides encoded thereby that have
30   been identified as having novel associations with a disease or pathology, or an abnormal state or condition, in a mammal. The present invention further identifies a set of proteins and polypeptides, including naturally occurring polypeptides, precursor forms or proproteins, or mature forms of the polypeptides or proteins, which are implicated as targets for therapeutic agents in the treatment of various diseases, pathologies, abnormal

17

states and conditions. A target may be employed in any of a variety of screening methodologies in order to identify candidate therapeutic agents which interact with the target and in so doing exert a desired or favorable effect. The candidate therapeutic agent is identified by screening a large collection of substances or compounds in an important

5   embodiment of the invention. Such a collection may comprise a combinatorial library of substances or compounds in which, in at least one subset of substances or compounds, the individual members are related to each other by simple structural variations based on a particular canonical or basic chemical structure. The variations may include, by way of nonlimiting example, changes in length or identity of a basic framework of bonded atoms;

10   changes in number, composition and disposition of ringed structures, bridge structures, alicyclic rings, and aromatic rings; and changes in pendent or substituents atoms or groups that are bonded at particular positions to the basic framework of bonded atoms or to the ringed structures, the bridge structures, the alicyclic structures, or the aromatic structures.

A polypeptide or protein described herein, and that serves as a target in the

15   screening procedure, includes the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, e.g., the full-length gene product, encoded by the corresponding gene. The naturally occurring polypeptide also includes the polypeptide, precursor or proprotein encoded by an open reading frame described herein. A "mature" form of a polypeptide or protein arises as

20   a result of one or more naturally occurring processing steps as they may occur within the cell, including a host cell. The processing steps occur as the gene product arises, e.g., via cleavage of the amino-terminal methionine residue encoded by the initiation codon of an open reading frame, or the proteolytic cleavage of a signal peptide or leader sequence. Thus, a mature form arising from a precursor polypeptide or protein that has residues 1 to

25   N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an amino-terminal signal sequence from residue 1 to residue M is cleaved, includes the residues from residue M+1 to residue N remaining. A "mature" form of a polypeptide or protein may also arise from non-proteolytic post-

30   translational modification. Such non-proteolytic processes include, e.g., glycosylation, myristylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or the combination of any of them.

As used herein, "identical" residues correspond to those residues in a comparison between two sequences where the equivalent nucleotide base or amino acid residue in an

18

alignment of two sequences is the same residue. Residues are alternatively described as "similar" or "positive" when the comparisons between two sequences in an alignment show that residues in an equivalent position in a comparison are either the same amino acid or a conserved amino acid as defined below.

5   As used herein, a "chemical composition" relates to a composition including at least one compound that is either synthesized or extracted from a natural source. A chemical compound may be the product of a defined synthetic procedure. Such a synthesized compound is understood herein to have defined properties in terms of molecular formula, molecular structure relating the association of bonded atoms to each other, physical

10   properties such as chromatographic or spectroscopic characterizations, and the like. A compound extracted from a natural source is advantageously analyzed by chemical and physical methods in order to provide a representation of its defined properties, including its molecular formula, molecular structure relating the association of bonded atoms to each other, physical properties such as chromatographic or spectroscopic characterizations, and

15   the like.

        As used herein, a "candidate therapeutic agent" is a chemical compound that includes at least one substance shown to bind to a target biopolymer. In important embodiments of the invention, the target biopolymer is a protein or polypeptide, a nucleic acid, a polysaccharide or proteoglycan, or a lipid such as a complex lipid. The method of

20   identifying compounds that bind to the target effectively eliminates compounds with little or no binding affinity, thereby increasing the potential that the identified chemical compound may have beneficial therapeutic applications. In cases where the "candidate therapeutic agent" is a mixture of more than one chemical compound, subsequent screening procedures may be carried out to identify the particular substance in the mixture that is the

25   binding compound, and that is to be identified as a candidate therapeutic agent.

        As used herein, a "pharmaceutical agent" is provided by screening a candidate therapeutic agent using models for a disease state or pathology in order to identify a candidate exerting a desired or beneficial therapeutic effect with relation to the disease or pathology. Such a candidate that successfully provides such an effect is termed a

30   pharmaceutical agent herein. Nonlimiting examples of model systems that may be used in such screens include particular cell lines, cultured cells, tissue preparations, whole tissues, organ preparations, intact organs, and nonhuman mammals. Screens employing at least one system, and preferably more than one system, may be employed in order to identify a

19

pharmaceutical agent. Any pharmaceutical agent so identified may be pursued in further investigation using human subjects.

NOVX Nucleic Acids and Polypeptides

5

NOVX clones

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence

10   of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy.

15   Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

20   The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a

25   small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide

30   comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 178; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 178, wherein any amino acid in the mature form is changed to a different

35   amino acid, provided that no more than 15% of the amino acid residues in the sequence of

20

the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 178; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 178 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 178; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 178 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 178; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 178, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 178 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO:

21

2n-1, wherein n is an integer between 1 and 178; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (e.g., NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA), RNA molecules (e.g., mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

An NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional

22

processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, e.g., 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (e.g., brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the invention, e.g., a nucleic acid molecule having the nucleotide sequence SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178 as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, et al., (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, et al., (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

23

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

5   Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a

10   genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length

15   would further comprise at least 6 contiguous nucleotides SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide from the group

20   consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, or a portion of this nucleotide sequence (e.g., a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of an NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178 is one that is sufficiently

25   complementary to the nucleotide sequence from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178 that it can hydrogen bond with little or no mismatches to the nucleotide sequence from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen

30   base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of

24

another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5' direction of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See e.g.

25

Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or

5   amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an NOVX

10   polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence

15   encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

20   An NOVX polypeptide is encoded by the open reading frame ("ORF") of an NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop"

25   codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

30   The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide

26

sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178; or an anti-sense strand nucleotide sequence of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178.

5   Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues

10   which mis-express an NOVX protein, such as by measuring a level of an NOVX-encoding nucleic acid in a sample of cells from a subject e.g., detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

        "A polypeptide having a biologically-active portion of an NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity

15   of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, that encodes a polypeptide having an NOVX biological activity (the biological activities of the NOVX proteins are described

20   below), expressing the encoded portion of NOVX protein (e.g., by recombinant expression in vitro) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

25   The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178 due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178. In another embodiment, an isolated nucleic acid molecule of

30   the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO: 2n, wherein n is an integer between 1 and 178.

        In addition to the human NOVX nucleotide sequences shown in SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the

35   NOVX polypeptides may exist within a population (e.g., the human population). Such

27

genetic polymorphism in the NOVX genes may exist among individuals within a
population due to natural allelic variation. As used herein, the terms "gene" and
"recombinant gene" refer to nucleic acid molecules comprising an open reading frame
(ORF) encoding an NOVX protein, preferably a vertebrate NOVX protein. Such natural
allelic variations can typically result in 1-5% variance in the nucleotide sequence of the
NOVX genes. Any and all such nucleotide variations and resulting amino acid
polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation
and that do not alter the functional activity of the NOVX polypeptides, are intended to be
within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and
thus that have a nucleotide sequence that differs from the human SEQ ID NO: 2n-1,
wherein n is an integer between 1 and 178 are intended to be within the scope of the
invention. Nucleic acid molecules corresponding to natural allelic variants and
homologues of the NOVX cDNAs of the invention can be isolated based on their
homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or
a portion thereof, as a hybridization probe according to standard hybridization techniques
under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the
invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the
nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 2n-1, wherein n
is an integer between 1 and 178. In another embodiment, the nucleic acid is at least 10, 25,
50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another
embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding
region. As used herein, the term "hybridizes under stringent conditions" is intended to
describe conditions for hybridization and washing under which nucleotide sequences at
least 60% homologous to each other typically remain hybridized to each other.

Homologs (i.e., nucleic acids encoding NOVX proteins derived from species other
than human) or other related sequences (e.g., paralogs) can be obtained by low, moderate
or high stringency hybridization with all or a portion of the particular human sequence as a
probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions
under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to
no other sequences. Stringent conditions are sequence-dependent and will be different in
different circumstances. Longer sequences hybridize specifically at higher temperatures

28

than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize

5   to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at
pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or

10   oligonucleotides (e.g., 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, et al., (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons,

15   N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm

20   DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence

25   that occurs in nature (e.g., encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency

30   hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. See, e.g., Ausubel, et al. (eds.), 1993, CURRENT PROTOCOLS IN

MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NO: 2n-1, wherein n is an integer
5   between 1 and 178, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM
10   Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (e.g., as employed for cross-species hybridizations). See, e.g., Ausubel, et al. (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981.
15   *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may
20   exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, thereby leading to changes in the amino acid sequences of the encoded NOVX proteins, without altering the functional ability of said NOVX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential"
25   amino acid residues can be made in the sequence SEQ ID NO: 2n, wherein n is an integer between 1 and 178. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the
30   invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO: 2n-1, wherein n is an

integer between 1 and 178 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences SEQ ID NO: 2n, wherein n is an integer between 1 and 178. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO: 2n, wherein n is an integer between 1 and 178; more preferably at least about 70% homologous SEQ ID NO: 2n, wherein n is an integer between 1 and 178; still more preferably at least about 80% homologous to SEQ ID NO: 2n, wherein n is an integer between 1 and 178; even more preferably at least about 90% homologous to SEQ ID NO: 2n, wherein n is an integer between 1 and 178; and most preferably at least about 95% homologous to SEQ ID NO: 2n, wherein n is an integer between 1 and 178.

An isolated nucleic acid molecule encoding an NOVX protein homologous to the protein of SEQ ID NO: 2n, wherein n is an integer between 1 and 178 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178 standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis

31

SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain
5   interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may
10   be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (*i*) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or
15   biologically-active portions thereof, (*ii*) complex formation between a mutant NOVX protein and an NOVX ligand; or (*iii*) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).
20

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the
25   nucleotide sequence of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid
30   molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of an NOVX protein of SEQ ID NO: 2n, wherein n is an integer between 1 and 178, or antisense nucleic acids complementary to an NOVX nucleic acid sequence of SEQ ID NO:
35   2n-1, wherein n is an integer between 1 and 178, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense

5   nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein,
10   antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the
15   region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically
20   synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic
25   acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine,
30   2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil,

33

5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense

5   orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or

10   genomic DNA encoding an NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of

15   administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense

20   nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

25   In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. *See, e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a

30   2'-o-methylribonucleotide (*See, e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330.

34

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the
5   modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a
10   complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for an NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of an NOVX cDNA disclosed herein (*i.e.*, SEQ ID NO: 2n-1, wherein
15   n is an integer between 1 and 178). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease
20   activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent
25   transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization,
30   or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural

35

nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc.*

5   *Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of

10   single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S_1 nucleases (*See,* Hyrup, *et al.,* 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*See,* Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.,* 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their

15   stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the

20   DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see,* Hyrup, et al., 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.,* 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For

25   example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.,* Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule

30   with a 5' PNA segment and a 3' DNA segment. *See, e.g.,* Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.,* Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as

peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO

5   89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the

10   like.

NOVX Polypeptides

         A polypeptide according to the invention includes a polypeptide including the
15   amino acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID
NO: 2n, wherein n is an integer between 1 and 178. The invention also includes a mutant
or variant protein any of whose residues may be changed from the corresponding residues
shown in SEQ ID NO: 2n, wherein n is an integer between 1 and 178 while still encoding a
protein that maintains its NOVX activities and physiological functions, or a functional
20   fragment thereof.

         In general, an NOVX variant that preserves NOVX-like function includes any
variant in which residues at a particular position in the sequence have been substituted by
other amino acids, and further include the possibility of inserting an additional residue or
residues between two residues of the parent protein as well as the possibility of deleting
25   one or more residues from the parent sequence. Any amino acid substitution, insertion, or
deletion is encompassed by the invention. In favorable circumstances, the substitution is a
conservative substitution as defined above.

         One aspect of the invention pertains to isolated NOVX proteins, and biologically-
active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also
30   provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX
antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue
sources by an appropriate purification scheme using standard protein purification
techniques. In another embodiment, NOVX proteins are produced by recombinant DNA
techniques. Alternative to recombinant expression, an NOVX protein or polypeptide can
35   be synthesized chemically using standard peptide synthesis techniques.

37

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language

5   "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more

10   preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably

15   less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals"

20   includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

25   Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (e.g., the amino acid sequence shown in SEQ ID NO: 2n, wherein n is an integer between 1 and 178) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of an NOVX protein. Typically, biologically-

30   active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of an NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NO: 2n, wherein n is an integer between 1 and 178. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO: 2n, wherein n is an integer between 1 and 178, and retains the functional activity of the protein of SEQ ID NO: 2n, wherein n is an integer between 1 and 178, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NO: 2n, wherein n is an integer between 1 and 178, and retains the functional activity of the NOVX proteins of SEQ ID NO: 2n, wherein n is an integer between 1 and 178.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See,* Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178.

39

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of

5   positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a

10   characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

15   **Chimeric and Fusion Proteins**

The invention also provides NOVX chimeric or fusion proteins. As used herein, an NOVX "chimeric protein" or "fusion protein" comprises an NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a

20   polypeptide having an amino acid sequence corresponding to an NOVX protein SEQ ID NO: 2n, wherein n is an integer between 1 and 178, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, e.g., a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within an

25   NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of an NOVX protein. In one embodiment, an NOVX fusion protein comprises at least one biologically-active portion of an NOVX protein. In another embodiment, an NOVX fusion protein comprises at least two biologically-active portions of an NOVX protein. In yet another embodiment, an NOVX fusion protein comprises at least three biologically-active

30   portions of an NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the

35   NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase)

40

sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is an NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells),

5   expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the

10   invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between an NOVX ligand and an NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction in vivo. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of an NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful

15   therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (e.g. promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with an NOVX ligand.

20   An NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, e.g., by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as

25   appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and

30   reamplified to generate a chimeric gene sequence (see, e.g., Ausubel, et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

41

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either

5   NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally

10   occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side

15   effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or

20   antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as

25   individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an

30   appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be
used to generate a variegated population of NOVX fragments for screening and subsequent
selection of variants of an NOVX protein. In one embodiment, a library of coding
sequence fragments can be generated by treating a double stranded PCR fragment of an
NOVX coding sequence with a nuclease under conditions wherein nicking occurs only
about once per molecule, denaturing the double stranded DNA, renaturing the DNA to
form double-stranded DNA that can include sense/antisense pairs from different nicked
products, removing single stranded portions from reformed duplexes by treatment with S_1
nuclease, and ligating the resulting fragment library into an expression vector. By this
method, expression libraries can be derived which encodes N-terminal and internal
fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of
combinatorial libraries made by point mutations or truncation, and for screening cDNA
libraries for gene products having a selected property. Such techniques are adaptable for
rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX
proteins. The most widely used techniques, which are amenable to high throughput
analysis, for screening large gene libraries typically include cloning the gene library into
replicable expression vectors, transforming appropriate cells with the resulting library of
vectors, and expressing the combinatorial genes under conditions in which detection of a
desired activity facilitates isolation of the vector encoding the gene whose product was
detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the
frequency of functional mutants in the libraries, can be used in combination with the
screening assays to identify NOVX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc.
Natl. Acad. Sci. USA* **89**: 7811-7815; Delgrave, *et al.,* 1993. *Protein Engineering*
6:327-331.

NOVX Antibodies

The term "antibody" as used herein refers to immunoglobulin molecules and
immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that
contain an antigen binding site that specifically binds (immunoreacts with) an antigen.
Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single

43

chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in

5   humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that

10   immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO:

15   2n, wherein n is an integer between 1 and 178, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes

20   encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, e.g., a hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will

25   indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier

30   transformation. See, e.g., Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety. Antibodies that are specific for one or more domains within an

44

antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that

5   immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor

10   Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g.,

15   rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may

20   be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete

25   and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose

30   dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide

45

primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D.

5   Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as

10   used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the

15   antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies

20   that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human

25   mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell

30   lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium

46

for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a

5   medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J.

10   Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is

15   determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic

20   applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding,1986). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium

25   and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel

30   electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding

47

the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce

5   immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the

10   coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

15   **Humanized Antibodies**

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric

20   immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')$_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature,

25   332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor

30   in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will

48

comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

5   **Human Antibodies**

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma

10   technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983.

15   Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991);

20   Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire.

25   This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al,(Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93

30   (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light

49

immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S.

5   Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art

10   including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

15   **Bispecific Antibodies**

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or

20   receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random

25   assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659

30   (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least

part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and

5   are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at

10   least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or

15   threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')2 bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific

20   antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')2 fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB)

25   derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically

30   coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')2 molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well

as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies

5 have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody

10 homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains

15 of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific

20 antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3,

25 CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA,

30 DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted

5   cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include

10 iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector

15 function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et

20 al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al.,

25 Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an

30 enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A

54

chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the

5   tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as

10   dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin

15   can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such

20   streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

25   **Immunoliposomes**

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545.

30   Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the

55

antibody of the present invention can be conjugated to the liposomes as described in Martin et al ., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon et al., J. National Cancer Inst., 81(19): 1484 (1989).

5

Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (e.g., for use in measuring levels of the protein within appropriate physiological samples, for use in

10 diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds (see below).

An antibody specific for a protein of the invention can be used to isolate the protein

15 by standard techniques, such as immunoaffinity chromatography or immunoprecipitation. Such an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein.

20 Antibodies directed against the protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials,

25 bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride

30 or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

56

Antibody Therapeutics

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and

57

considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends,

5   Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the

10   smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA, 90: 7889-7893

15   (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such

20   molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate)

25   microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

30   Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S.

58

Pat. No. 3,773,919), copolymers of L-glutamic acid and γ ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT ^{TM} (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers

5   such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an

10   analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., F_{ab} or $F_{(ab)2}$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the

15   probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells

20   and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of an analyte mRNA include

25   Northern hybridizations and in situ hybridizations. In vitro techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. In vitro techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in

30   Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulus, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Thory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, in vivo techniques for detection of an analyte protein include introducing into a subject a labeled anti-an analyte protein antibody.

For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory

5 sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired,

10 etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression

15 of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be

20 transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein

25 encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion

30 moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5

61

(Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION

5   TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.*, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN

10   ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.*, Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis

15   techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen

20   Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.). Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

25   In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For

30   example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific

5   promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament

10   promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249:

15   374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that

20   allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or

25   cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see*, *e.g.*,

30   Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms

63

refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

5   A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

           Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional
10   transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found
15   in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

           For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may
20   integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell
25   on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

           A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture,
30   can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable

64

medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

5

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in

10   which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more

15   preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an

20   encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to

25   development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (e.g., by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences SEQ ID NO:

30   2n-1, wherein n is an integer between 1 and 178 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the

35   efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be

65

operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and

5 Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional

10 animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an NOVX gene into which a deletion, addition or substitution has been

15 introduced to thereby alter, e.g., functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (e.g., the cDNA of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178 can be used to construct a homologous

20 recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous

25 recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur

30 between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. See, e.g., Thomas, et al., 1987. Cell 51: 503 for a description of homologous

recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

5   The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-

10   recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968;

15   and WO 93/04169.

      In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci.*

20   *USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic

25   animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

      Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit

30   the growth cycle and enter G$_0$ phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster

animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

5

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the

10 nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical

15 Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any

20 conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include

25 parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl

30 alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or

35 multiple dose vials made of glass or plastic.

68

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF,

5   Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example,

10 glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens,

15 chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

20   Sterile injectable solutions can be prepared by incorporating the active compound (e.g., an NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required

25 other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

        Oral compositions generally include an inert diluent or an edible carrier. They can

30 be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible

binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid,

5   Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant,

10   e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic

15   acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention

20   enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides,

25   polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be

30   prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be

70

treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular

5   therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see, e.g., U.S. Patent No. 5,328,470) or by

10   stereotactic injection (see, e.g., Chen, et al., 1994. Proc. Natl. Acad. Sci. USA 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the

15   pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

20   **Screening and Detection Methods**

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (e.g., in a biological sample) or a genetic lesion in

25   an NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (e.g.; diabetes

30   (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease(possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins

35   and modulate NOVX activity. In yet a further aspect, the invention can be used in methods

71

to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

5

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides,

10   peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test

15   compounds which bind to or modulate the activity of the membrane-bound form of an NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring

20   deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

25   A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can

30   be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed.*

72

Engl. 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to an NOVX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or

73

biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule. As used herein, a "target molecule" is a molecule

5 with which an NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses an NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. An NOVX target molecule can be a non-NOVX molecule or an NOVX protein or polypeptide of the invention. In one

10 embodiment, an NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (e.g. a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with

15 NOVX.

Determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule can be accomplished by

20 determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (i.e. intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising an NOVX-responsive regulatory element operatively linked to a nucleic acid

25 encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting an NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or

30 biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to

74

interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

5   In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by

10   determining the ability of the NOVX protein to bind to an NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate an NOVX target molecule. For example, the catalytic/enzymatic activity

15   of the target molecule on an appropriate substrate can be determined as described, *supra*.

      In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein

20   determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of an NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the

25   membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®,

30   Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylammminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylammminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

      In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate

75

separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the

5   reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter

10   plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined

15   either directly or indirectly, for example, as described, supra. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target

20   molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or

25   target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target

30   molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX

76

mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA

5 or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA

10 or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268:

15 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the

20 NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*,

25 GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This

30 proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

5

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective genes on a chromosome; and, thus, locate gene regions associated

10   with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

15   Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences, SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a

20   chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX, sequences can be used to rapidly select primers that do not span more than one

25   exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals

30   (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of

35   hybrid cell lines can be established. Each cell line in a panel contains either a single human

78

chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.,* D'Eustachio, *et al.,* 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human

5   chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific

10   chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes

15   can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection.

20   Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see,* Verma, *et al.,* HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for

25   marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical

30   position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.,* in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through

linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.,* Egeland, *et al.,* 1987. *Nature,* 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a

5   mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence.

10   Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from

15   minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

20   Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

25   Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic

30   variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide

5   positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

10   **Predictive Medicine**

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining

15   NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia,

20   cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of

25   developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in an NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

30   Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the

81

genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

5   These and other agents are described in further detail in the following sections.

Diagnostic Assays

          An exemplary method for detecting the presence or absence of NOVX in a

10   biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX

15   mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 178, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the

20   diagnostic assays of the invention are described herein.

          An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')$_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended

25   to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with

30   fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern

35   hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX

82

protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX

5   antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is

10   a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and

15   comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for

20   determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

25   **Prognostic Assays**

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the

30   preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression

35   or activity in which a test sample is obtained from a subject and NOVX protein or nucleic

83

acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample

5   can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For

10   example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic

15   acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in an NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various

20   embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding an NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from an NOVX gene;

25   (*ii*) an addition of one or more nucleotides to an NOVX gene; (*iii*) a substitution of one or more nucleotides of an NOVX gene, (*iv*) a chromosomal rearrangement of an NOVX gene; (*v*) an alteration in the level of a messenger RNA transcript of an NOVX gene, (*vi*) aberrant modification of an NOVX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of an

30   NOVX gene, (*viii*) a non-wild-type level of an NOVX protein, (*ix*) allelic loss of an NOVX gene, and (*x*) inappropriate post-translational modification of an NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in an NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and

5 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells

10 from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to an NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the

15 length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional

20 amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al,* 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low

25 numbers.

In an alternative embodiment, mutations in an NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and

30 compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.*, Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic

5 mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is

10 followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the

15 art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing

20 procedures can be utilized when performing the diagnostic assays (*see, e.g.*, Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in

25 which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.*, Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded

30 duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with

86

hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.,* 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.,* 1992. *Methods*

5   *Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations

10   in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.,* 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on an NOVX sequence, *e.g.,* a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test

15   cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation

20   polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of

25   single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes

30   heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.,* 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.,* 1985. *Nature* 313: 495.

When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and

5   sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions

10   that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.,* 1986. *Nature* 324: 163; Saiki, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

15   Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.,* 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the

20   extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.,* 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be

25   performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing

30   pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

5

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (e.g., NOVX gene expression), as identified by a screening assay described herein

10 can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X

15 and wasting disorders associated with chronic diseases and various cancers.) In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood

20 concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid,

25 or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997.

30 *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example,

35 glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited

89

enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome PREGNANCY ZONE PROTEIN PRECURSOR enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification. Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of NOVX (e.g., the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials.

90

For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an

5   agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as

10   a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) that modulates NOVX activity (e.g., identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for

15   example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity

20   of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the

25   effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of an NOVX protein, mRNA, or genomic DNA in the

30   preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or

91

samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease

5   expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

10   The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic

15   stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of

20   Albright Hereditary Ostoeodystrophy, and other diseases, disorders and conditions of the like.

These methods of treatment will be discussed more fully, below.

Disease and Disorders

25

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that

30   may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are

35   utilized to "knockout" endogenous function of an aforementioned peptide by homologous

92

recombination (*see, e.g.,* Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

5   Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs,

10   derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.,* from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are

15   not limited to, immunoassays (*e.g.,* by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.,* Northern assays, dot blots, *in situ* hybridization, and the like).

20   **Prophylactic Methods**

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one

25   NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression.

30   Depending upon the type of NOVX aberrancy, for example, an NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of

5   NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of an NOVX protein, a peptide, an NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX

10   protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As

15   such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of an NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another

20   embodiment, the method involves administering an NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situ*ations in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder

25   characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic
30

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative

35   cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic

94

exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to

5   administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The NOVX nucleic acids and proteins of the invention are useful in potential

10   prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting

15   disorders associated with chronic diseases and various cancers.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity,

20   infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the

25   presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.,* some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

30
Sequence Analyses

The sequence of NOVX was derived by laboratory cloning of cDNA fragments, by *in silico* prediction of the sequence. cDNA fragments covering either the full length of the

35   DNA sequence, or part of the sequence, or both, were cloned. *In silico* prediction was

95

based on sequences available in CuraGen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

5   The laboratory cloning was performed using one or more of the methods summarized below:

SeqCalling™Technology: cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and
10   developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen Corporation's SeqCalling technology which is disclosed in full in U. S. Ser. Nos.
15   09/417,386 filed Oct. 13, 1999, and 09/614,505 filed July 11, 2000. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatics programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as
20   components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Variant sequences are also included in this application. A variant sequence can
25   include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide
30   or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the

genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression,

5   and stability of transcribed message.

Presented information includes that associated with genomic clones, public genes and ESTs sharing sequence identity with the disclosed sequence and CuraGen Corporation's Electronic Northern bioinformatic tool.

10   Examples

Example A: Sequence related information

The NOV1 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 1A.

Table 1A. NOV1 Sequence Analysis		
	SEQ ID NO: 1	711 bp
NOV1a, CG58522-01 DNA Sequence	TGCAGAATGAACCAAGGAGACTCAAACCCAGCAGCTACTCCGCATGCGGCAGAAGACA TTCAAGGAGATGACAGATGGATGTGTCAGCACAACAGATTTGTTTTGGACTGTAAAGA CAAACAGCCTGATGTACCATTTGCGGGAGGCTCCGTGGTGCAGTTACTGCAGCCATAT GAGATATGGCGAGAGCTTTTTTCCCCACTTCATGCACTGAATTTTGGAACTGGGGGAG ATACAACAAGACATGTTTTGTGGAGACTAAAGAGCGGAGAACTGGGGAATACTAAGCC TAAGGTCATTGTTTTCTGGCTAGGAAGAAACAACCATGAAAATATGGCAGAAGAGGTA GCAGGTGGTATGGCGGCCATCGTACAACTTATCAACACAAGGCAGCCACAGGCCAAAA TCATTGTATTTGATCTGTTACCTCAAGGTGAGAAACCCAACCCTTTGAGGCAAAAGAA CGCCAAGGTGAACCCACTCGTCAAGATTTCGCTGCTGAAACTTACCAACGTGCAGCTC CTGGATACTGACAGGGGTTTCGTGCACTCCGACCGTGCCATCTCCTGCCACGACATGT TTGATTTTCTGCATTTGACAGGAGGTGGCTACTCAAAGGTCTGCAAACCCTTGAATGA ACTGATCATGCAGTTGTTGGAGGAAACACCTGAGGAGAAACAAACCACCATTGCCTGA CTGGCTCCCATGAGT	
	ORF Start: ATG at 7	ORF Stop: TGA at 694
	SEQ ID NO: 2	229 aa	MW at 25656.2kD
NOV1a, CG58522-01 Protein Sequence	MNQGDSNPAATPHAAEDIQGDDRWMCQHNRFVLDCKDKQPDVPFAGGSVVQLLQPYEI WRELFSPLHALNFGTGGDTTRHVLWRLKSGELGNTKPKVIVFWLGRNNHENMAEEVAG GMAAIVQLINTRQPQAKIIVFDLLPQGEKPNPLRQKNAKVNPLVKISLLKLTNVQLLD TDRGFVHSDRAISCHDMFDFLHLTGGGYSKVCKPLNELIMQLLEETPEEKQTTIA	

Further analysis of the NOV1a protein yielded the following properties shown in

15   Table 1B.

Table 1B. Protein Sequence Properties NOV1a	
Psort analysis:	0.6500 probability located in cytoplasm; 0.2340 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space; 0.0000 probability located in endoplasmic reticulum (membrane)

97

SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 1C.

Table 1C. Geneseq Results for NOV1a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB49433	Human beta platelet activating factor acetylhydrolase - Homo sapiens, 229 aa. [US6146868-A, 14-NOV-2000]	1..229 1..229	196/229 (85%) 209/229 (90%)	e-114
AAB49432	Rat beta platelet activating factor acetylhydrolase - Rattus norvegicus, 229 aa. [US6146868-A, 14-NOV-2000]	1..229 1..229	195/229 (85%) 208/229 (90%)	e-114
AAB49434	Murine beta platelet activating factor acetylhydrolase - Mus musculus, 229 aa. [US6146868-A, 14-NOV-2000]	1..229 1..229	192/229 (83%) 205/229 (88%)	e-111
AAB49436	Bovine gamma platelet activating factor acetylhydrolase - Bos taurus, 232 aa. [US6146868-A, 14-NOV-2000]	4..219 3..218	124/216 (57%) 165/216 (75%)	5e-74
AAB49435	Human gamma platelet activating factor acetylhydrolase - Homo sapiens, 231 aa. [US6146868-A, 14-NOV-2000]	4..219 3..218	124/216 (57%) 164/216 (75%)	2e-73

In a BLAST search of public sequence databases, the NOV1a protein was found to have homology to the proteins shown in the BLASTP data in Table 1D.

Table 1D. Public BLASTP Results for NOV1a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q29459	Platelet-activating factor acetylhydrolase IB beta subunit (EC 3.1.1.47) (PAF acetylhydrolase 30 kDa subunit) (PAF-AH 30 kDa subunit) (PAF-AH beta subunit) (PAFAH beta subunit) - Homo sapiens (Human), and, 229 aa.	1..229 1..229	196/229 (85%) 209/229 (90%)	e-114
O35264	Platelet-activating factor acetylhydrolase IB beta subunit (EC 3.1.1.47) (PAF acetylhydrolase 30 kDa subunit) (PAF-AH 30 kDa subunit) (PAF-AH beta subunit) (PAFAH beta subunit) (Platelet-activating factor acetylhydrolase alpha 2 subunit) (PAF-AH alpha 2) - Rattus norvegicus (Rat), 229 aa.	1..229 1..229	195/229 (85%) 208/229 (90%)	e-113
Q61206	Platelet-activating factor acetylhydrolase IB beta subunit (EC 3.1.1.47) (PAF acetylhydrolase 30 kDa subunit) (PAF-AH 30 kDa subunit) (PAF-AH beta subunit) (PAFAH beta subunit) - Mus musculus (Mouse), 229 aa.	1..229 1..229	192/229 (83%) 205/229 (88%)	e-111
Q29460	Platelet-activating factor acetylhydrolase IB gamma subunit (EC 3.1.1.47) (PAF acetylhydrolase 29 kDa subunit) (PAF-AH 29 kDa subunit) (PAF-AH gamma subunit) (PAFAH gamma subunit) - Bos taurus (Bovine), 232 aa.	4..219 3..218	125/216 (57%) 165/216 (75%)	8e-74
Q15102	Platelet-activating factor acetylhydrolase IB gamma subunit (EC 3.1.1.47) (PAF acetylhydrolase 29 kDa subunit) (PAF-AH 29 kDa subunit) (PAF-AH gamma subunit) (PAFAH gamma subunit) - Homo sapiens (Human), 231 aa.	4..219 3..218	124/216 (57%) 164/216 (75%)	7e-73

PFam analysis predicts that the NOV1a protein contains the domains shown in the Table 1E.

Table 1E. Domain Analysis of NOV1a			
Pfam Domain	NOV1a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PAF-AH: domain 1 of 1	7..221	150/215 (70%) 186/215 (87%)	6e-147

Example 2.

The NOV2 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 2A.

Table 2A. NOV2 Sequence Analysis		
	SEQ ID NO: 3	1457 bp
NOV2a, CG58520-01 DNA Sequence	CGATTCCGATGGGTCCTTTGAAAGCTTTTCTCTTCTCCCCTTTTCTTCTGCGGAGTCA AAGTAGAGGGGTGAGGTTGGTCTTCTTGTTACTGACCCTGCATTTGGGAAACGTTGAT AAGGCAGATGATGAAGATGATGAGGATTTAACGGTGAACAAAACCTGGGTCTTGGCCC CAAAAATTCATGAAGGAGATATCACACAAATTCTGAATTCATTGCTTCAAGGCTATGA CAATAAACTTCGTCCAGATATAGGAGTGAGGCCCACAGTAATTGAAACTGATGTTTAT GTAAACAGCATTGGACCAGTTGATCCAATTAATATGGAATATACAATAGATATAATTT TTGCCCAAACCTGGTTTGACAGTCGTTTAAAATTCAATAGTACCATGAAAGTGCTTAT GCTTAACAGTAATATGGTTGGAAAAATTTGGATTCCTGACACTTTCTTCAGAAACTCA AGAAAATCTGATGCTCACTGGATAACAACTCCTAATCGTCTGCTTCGAATTTGGAATG ATGGACGAGTTCTGTATACTCTAAGGAGATTGACAATTAATGCAGAATGTTATCTTCA GCTTCATAACTTTCCCATGGATGAACATTCCTGTCCACTGGAATTTTCAAGCTTCTCT ATAGATGGATACCCTAAAAATGAAATTGAGTTATCAATGGAAGCGAAGTTCTGTGGAA GTGGGCGACACAAGATCCGGAGATTATATCAGTTTGCATTTGTAGGGTTACGGAACTC AACTGAAATCACTCACACGATCTCTGGGGATTATGTTATCATGACAATTTTTTTTGAC CTGAGCAGAAGAATGGGATATTTCACTATTCAGACCTACATTCCATGCATTCTGACAG TTGTTCTTTCTTGGGTGTCTTTTTGGATCAATAAAGATGCAGTGCCTGCAAGAACATC GTTGGGTATGACATCTATAGGTATCACTACAGTTCTGACTATGACAACCCTGAGTACA ATTGCCAGGAAGTCTTTACCTAAGGTTTCTTATGTGACTGCGATGGATCTCTTTGTTT CTGTTTGTTTCATTTTTGTTTTTGCAGCCTTGATGGAATATGGAACCTTGCATTATTT TACCAGCAACCAAAAAGGAAAGACTGCTACTAAAGACAGAAAGCTAAAAAATAAAGCC TCGACTCCTGGTCTCCATCCTGGATCCACTCTGATTCCAATGAATAATATTTCTGTGC CGCAAGAAGATGATTATGGGTATCAGTGTTTGGAGGGCAAAGATTGTGCCAGCTTCTT CTGTTGCTTTGAAGACTGCAGAACAGGATCTTGGAGGGAAGGAAGGATACACATACGC ATTGCCAAAATTGACTCTTATTCTAGAATATTTTTCCCAACCGCTTTTGCCCTGTTCA ACTTGGTTTATTGGGTTGGCTATCTTTACTTAT**AAA**ATCTACTTCATAAGCAAAAATC AAAAGAA
	ORF Start: ATG at 9	ORF Stop: TAA at 1425
	SEQ ID NO: 4	472 aa · MW at 54100.9kD
NOV2a, CG58520-01 Protein Sequence	MGPLKAFLFSPFLLRSQSRGVRLVFLLLTLHLGNVDKADDEDDEDLTVNKTWVLAPKI HEGDITQILNSLLQGYDNKLRPDIGVRPTVIETDVYVNSIGPVDPINMEYTIDIIFAQ TWFDSRLKFNSTMKVLMLNSNMVGKIWIPDTFFRNSRKSDAHWITTPNRLLRIWNDGR VLYTLRRLTINAECYLQLHNFPMDEHSCPLEFSSFSIDGYPKNEIELSMEAKFCGSGR HKIRRLYQFAFVGLRNSTEITHTISGDYVIMTIFFDLSRRMGYFTIQTYIPCILTVVL SWVSFWINKDAVPARTSLGMTSIGITTVLTMTTLSTIARKSLPKVSYVTAMDLFVSVC FIFVFAALMEYGTLHYFTSNQKGKTATKDRKLKNKASTPGLHPGSTLIPMNNISVPQE DDYGYQCLEGKDCASFFCCFEDCRTGSWREGRIHIRIAKIDSYSRIFFPTAFALFNLV YWVGYLYL
	SEQ ID NO: 5	1521 bp
NOV2b, CG58520-02 DNA Sequence	CAACCAAGAGGCAAGAGGCGAGAGAAGGAAAAAAAAAAAAAGCGATGAGTTCGCCAAAT ATATGGAGCACAGGAAGCTCAGTCTACTCGACTCCTGTATTTTCACAGAAAATGACGG TGTGGATTCTGCTCCTGCTGTCGCTCTACCCTGGCTTCACTAGCCAGAAATCTGATGA TGACTATGAAGATTATGCTTCTAACAAAACATGGGTCTTGACTCCAAAAGTTCCTGAG GGTGATGTCACTGTCATCTTAAACAACCTGCTGGAAGGATATGACAATAAACTTCGGC CTGATATAGGAGTGAAGCCAACGTTAATTCACACAGACATGTATGTGAATAGCATTGG TCCAGTGAACGCTATCAATATGGAATACACTATTGATATATTTTTTGCGCAAACGTGG TATGACAGACGTTTGAAATTTAACAGCACCATTAAAGTCCTCCGATTGAACAGCAACA

	TGGTGGGGAAAATCTGGATTCCAGACACTTTCTTCAGAAATTCCAAAAAAGCTGATGC ACACTGGATCACCACCCCCAACAGGATGCTGAGAATTTGGAATGATGGTCGAGTGCTC TACACCCTAAGGTTGACAATTGATGCTGAGTGCCAATTACAATTGCACAACTTTCCAA TGGATGAACACTCCTGCCCCTTGGAGTTCTCCAGTTATGGCTATCCACGTGAAGAAAT TGTTTATCAATGGAAGCGAAGTTCTGTTGAAGTGGGCGACACAAGATCCTGGAGGCTT TATCAATTCTCATTTGTTGGTCTAAGAAATACCACCGAAGTAGTGAAGACAACTTCCG GAGATTATGTGGTCATGTCTGTCTACTTTGATCTGAGCAGAAGAATGGGATACTTTAC CATCCAGACCTATATCCCCTGCACACTCATTGTCGTCCTATCCTGGGTGTCTTCTGG ATCAATAAGGATGCTGTTCCAGCCAGAACATCTTTAGGTATCACCACTGTCCTGACAA TGACCACCCTCAGCACCATTGCCCGGAAATCGCTCCCCAAGGTCTCCTATGTCACAGC GATGGATCTCTTTGTATCTGTTTGTTTCATCTTTGTCTTCTCTGCTCTGGTGGAGTAT GGCACCTTGCATTATTTTGTCAGCAACCGGAAACCAAGCAAGGACAAAGATAAAAAGA AGAAAAACCCTCTTCTTCGGATGTTTTCCTTCAAGGCCCCTACCATTGATATCCGCCC AAGATCAGCAACCATTCAAATGAATAATGCTACACACCTTCAAGAGAGAGATGAAGAG TACGGCTATGAGTGTCTGGACGGCCAAGGACTGTGCCAGTTTTTTCTGCTGTTTTGAAG ATTGTCGAACAGGAGCTTGGAGACATGGGAGGATACATATCCGCATTGCCAAAATGGA CTCCTATGCTCGGATCTTCTTCCCCACTGCCTTCTGCCTGTTTAATCTGGTCTATTGG GTCTCCTACCTCTACCTGTGAGGAGGTATGGGTTTTACTGATATGGTTCTTATTCACT GAGTCTCATGGAG
	ORF Start: ATG at 44	ORF Stop: TGA at 1469
	SEQ ID NO: 6	475 aa	MW at 55184.9kD
NOV2b, CG58520-02 Protein Sequence	MSSPNIWSTGSSVYSTPVFSQKMTVWILLLLSLYPGFTSQKSDDDYEDYASNKTWVLT PKVPEGDVTVILNNLLEGYDNKLRPDIGVKPTLIHTDMYVNSIGPVNAINMEYTIDIF FAQTWYDRRLKFNSTIKVLRLNSNMVGKIWIPDTFFRNSKKADAHWITTPNRMLRIWN DGRVLYTLRLTIDAECQLQLHNFPMDEHSCPLEFSSYGYPREEIVYQWKRSSVEVGDT RSWRLYQFSFVGLRNTTEVVKTTSGDYVVMSVYFDLSRRMGYFTIQTYIPCTLIVVLS WVSFWINKDAVPARTSLGITTVLTMTTLSTIARKSLPKVSYVTAMDLFVSVCFIFVFS ALVEYGTLHYFVSNRKPSKDKDKKKKNPLLRMFSFKAPTIDIRPRSATIQMNNATHLQ ERDEEYGYECLDGKDCASFFCCFEDCRTGAWRHGRIHIRIAKMDSYARIFFPTAFCLF NLVYWVSYLYL
	SEQ ID NO: 7	1455 bp
NOV2c, CG58520-03 DNA Sequence	TAGTGCAGCACACGTAAAAAAGCGATTCCGATGGGTCCTTTGAAAGCTTTTCTCTTCT CCCCTTTTCTTCTGCGGAGTCAAAGTAGAGGGGTGAGGTTGGTCTTCTTGTTACTGAC CCTGCATTTGGGAAACTGGGTTGATAAGGCAGATGATGAAGATGATGAGGATTTAACG GTGAACAAAACCTGGGTCTTGGCCCCAAAAATTCATGAAGGAGATATCACACAAATTC TGAATTCATTGCTTCAAGGCTATGACAATAAACTTCGTCCAGATATAGGAGTGAGGCC CACAGTAATTGAAACTGATGTTTATGTAAACAGCATTGGACCAGTTGATCCAATTAAT ATGGAATATACAATAGATATAATTTTTGCCCAAACCTGGTTTGACAGTCGTTTAAAAT TCAATAGTACCATGAAAGTGCTTATGCTTAACAGTAATATGGTTGGAAAAATTTGGAT TCCTGACACTTTCTTCAGAAACTCAAGAAAATCTGATGCTCACTGGATAACAACTCCT AATCGTCTGCTTCGAATTTGGAATGATGGACGAGTTCTGTATACTCTAAGGTTGACAA TTAATGCAGAATGTTATCTTCAGCTTCATAACTTTCCCATGGATGAACATTCCTGTCC ACTGGAATTTTCAAGCGATGGATACCCTAAAAATGAAATTGAGTATAAGTGGAAAAAG CCCTCCGTAGAAGTGGCTGATCCTAAAATACTGGAGATTATATCAGTTTGCATTTGTAG GGTTACGGAACTCAACTGAAATCACTCACACGATCTCTGGTGATTATGTTATCATGAC AATTTTTTTTGACCTGAGCAGAAGAATGGGATATTTCACTATTCAGACCTACATTCCA TGCATTCTGACAGTTGTTCTTTCTTGGGTGTCTTTTTGGATCAATAAAGATGCAGTGC CTGCAAGAACATCGTTGGGTATCACTACAGTTCTGACTATGACAACCCTGAGTACAAT TGCCAGGAAGTCTTTACCTAAGGTTTCTTATGTGACTGCGATGGATCTCTTTGTTTCT GTTTGTTTCATTTTTTGTTTTTGCAGCCTTGATGGAATATGGAACCTTGCATTATTTTA CCAGCAACCAAAAAGGAAAGACTGCTACTAAAGACAGAAAGCTAAAAAATAAAGCCTC GGTAACTCCTGGTCTCCATCCTGGATCCACTCTGATTCCAATGAATAATATTTCTGTG CCGCAAGAAGATGATTATGGGTATCAGTGTTTGGAGGGCAAGGATTGTGCCAGCTTCT TCTGTTGCTTTGAAGACTGCAGAACAGGATCTTGGAGGGAAGGAAGGATACACATACG CATTGCCAAAATTGACTCTTATTCTAGAATATTTTTCCCAACCGCTTTTGCCCTGTTC AACTTGGTTTATTGGGTTGGCTATCTTTACTTATAAAATCTACTTCATAAGCAAAAAT CAAAA
	ORF Start: ATG at 31	ORF Stop: TAA at 1426
	SEQ ID NO: 8	465 aa	MW at 53597.3kD
NOV2c, CG58520-03 Protein Sequence	MGPLKAFLFSPFLLRSQSRGVRLVFLLLTLHLGNWVDKADDEDDEDLTVNKTWVLAPK IHEGDITQILNSLLQGYDNKLRPDIGVRPTVIETDVYVNSIGPVDPINMEYTIDIIFA QTWFDSRLKFNSTMKVLMLNSNMVGKIWIPDTFFRNSRKSDAHWITTPNRLLRIWNDG RVLYTLRLTINAECYLQLHNFPMDEHSCPLEFSSDGYPKNEIEYKWKKPSVEVADPKY WRLYQFAFVGLRNSTEITHTISGDYVIMTIFFDLSRRMGYFTIQTYIPCILTVVLSWV SFWINKDAVPARTSLGITTVLTMTTLSTIARKSLPKVSYVTAMDLFVSVCFIFVFAAL MEYGTLHYFTSNQKGKTATKDRKLKNKASVTPGLHPGSTLIPMNNISVPQEDDYGYQC LEGKDCASFFCCFEDCRTGSWREGRIHIRIAKIDSYSRIFFPTAFALFNLVYWVGYLY L

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 2B.

Table 2B. Comparison of NOV2a against NOV2b through NOV2c.		
Protein Sequence	**NOV2a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV2b	24..472 27..475	311/458 (67%) 352/458 (75%)
NOV2c	1..472 1..465	414/474 (87%) 415/474 (87%)

Further analysis of the NOV2a protein yielded the following properties shown in Table 2C.

Table 2C. Protein Sequence Properties NOV2a	
Psort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 38 and 39

5   A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 2D.

Table 2D. Geneseq Results for NOV2a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV2a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAM41007	Human polypeptide SEQ ID NO 5938 - Homo sapiens, 489 aa. [WO200153312-A1, 26-JUL-2001]	24..472 49..489	334/451 (74%) 379/451 (83%)	0.0
AAM39221	Human polypeptide SEQ ID NO 2366 - Homo sapiens, 467 aa. [WO200153312-A1, 26-JUL-2001]	24..472 27..467	334/451 (74%) 379/451 (83%)	0.0

AAR83968	GABA-A receptor gamma-3 subunit - Homo sapiens, 467 aa. [WO9529234-A1, 02-NOV-1995]	24..472 5..467	300/472 (63%) 356/472 (74%)	e-169
AAW59048	GABA-A receptor epsilon sub-unit related protein - Mammalia, 506 aa. [DE19644501-A1, 30-APR-1998]	62..472 70..506	193/448 (43%) 274/448 (61%)	e-102
AAW61045	Human GABA receptor epsilon subunit - Homo sapiens, 506 aa. [WO9823742-A1, 04-JUN-1998]	62..472 70..506	193/448 (43%) 274/448 (61%)	e-102

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2E.

Table 2E. Public BLASTP Results for NOV2a				
Protein Accession Number	Protein/Organism/Length	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P23574	Gamma-aminobutyric-acid receptor gamma-1 subunit precursor (GABA(A) receptor) - Rattus norvegicus (Rat), 465 aa.	1..472 1..465	426/475 (89%) 440/475 (91%)	0.0
Q9R0Y8	Gamma-aminobutyric-acid receptor gamma-1 subunit precursor (GABA(A) receptor) - Mus musculus (Mouse), 465 aa.	1..472 1..465	420/477 (88%) 434/477 (90%)	0.0
JH0824	gamma-aminobutyric acid A receptor gamma 1 chain precursor - chicken, 464 aa.	16..472 12..464	390/463 (84%) 416/463 (89%)	0.0
JH0316	gamma-aminobutyric acid A receptor gamma 2 chain alternatively spliced precursor - mouse, 466 aa.	24..472 26..466	336/451 (74%) 380/451 (83%)	0.0
P18508	Gamma-aminobutyric-acid receptor gamma-2 subunit precursor (GABA(A) receptor) - Rattus norvegicus (Rat), 466 aa.	24..472 26..466	335/451 (74%) 379/451 (83%)	0.0

PFam analysis predicts that the NOV2a protein contains the domains shown in the Table 2F.

Table 2F. Domain Analysis of NOV2a			
Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Neur_chan_LBD: domain 1 of 1	63..273	66/271 (24%) 162/271 (60%)	2.7e-56
Cys-protease-3C: domain 1 of 1	363..369	4/7 (57%) 6/7 (86%)	5.2
Neur_chan_memb: domain 1 of 1	280..466	44/297 (15%) 164/297 (55%)	1.2e-60

Example 3.

The NOV3 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 3A.

Table 3A. NOV3 Sequence Analysis		
	SEQ ID NO: 9	1440 bp
NOV3a, CG58518-01 DNA Sequence	GAAGAGATGGTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCTGGATCATATTGA AACCAAATGTTTGTGCTGCTTCTAACATCAAGATGACACACCAGCGGTGCTCCTCTTC AATGAAACAAACCTGGAAACAAGAAACTAGAATGAAGAAAGATGACAGTACCAAAGCG CGGCCTCAGAAATATGAGCAACTTCTCCATATAGAGGACAACGATTTCGCAATGAGAC CTGGATTTGGAGGTGAGTATTATCCTCTCAAAATTGGGTCTCCAGTGCCAGTAGGTAT AGATGTCCATGTTGAAAGCATTGACAGCATTTCAGAGACTAACATGGTAAGTTTCTTC ATGGGATATGACTTTACAATGACTTTTTATCTCAGGCATTACTGGAAAGACGAGAGGC TCTCCTTTCCTAGCACAGCAAACAAAAGCATGACATTTGATCATAGATTGACCAGAAA GATCTGGGTGCCTGATATCTTTTTTGTCCACTCTAAAAGATCCTTCATCCATGATACA ACTATGGAGAATATCATGCTGCGCGTACACCCTGATGGAAACGTCCTCCTAAGTCTCA GGAGGATAACGGTTTCGGCCATGTGCTTTATGGATTTCAGCAGGTTTCCTCTTGACAC TCAAAATTGTTCTCTTGAACTGGAAAGCGCCTACAATGAGGATGACCTAATGCTATAC TGGAAACACGGAAACAAGTCCTTAAATACTGAAGAACATATGTCCCTTTCTCAGTTCT TCATTGAAGACTTCAGTGCATCTAGTGGATTAGCTTTCTATAGCAGCACAGGTTGGTA CAATAGGCTTTTCATCAACTTTGTGCTAAGGAGGCATGTTTTCTTCTTTGTGCTGCAA ACCTATTTCCCAGCCATATTGATGGTGATGCTTTCATGGGTTTCATTTTGGATTGACC GAAGAGCTGTTCCTGCAAGAGTTTCCCTGGGTGGAATCACCACAGTGCTGACCATGTC CACAATCATCACTGCTGTGAGCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTG GATGTGTACCTGTGGGTCAGCTCCCTCTTTGTGTTCCTGTCAGTCATTGAGTATGCAG CTGTGAACTACCTCACCACAGTGGAAGAGCGGAAACAATTCAAGAAGACAGGAAAGGT ACAGATTTCTAGGATGTACAATATTGATGCAGTTCAAGCTATGGCCTTTGATGGTTGT TACCATGACAGCGAGATTGACATGGACCAGACTTCCCTCTCTCTAAACTCAGAAGACT TCATGAGAAGAAAATCGATATGCAGCCCCAGCACCGATTCATCTCGGATAAAGAGAAG AAAATCCCTAGGAGGACATGTTGGTAGAATCATTCTGGAAAACAACCATGTCATTGAC ACCTATTCTAGGATTTTATTCCCCATTGTGTATATCTTTATTTAATTT
	ORF Start: ATG at 7	ORF Stop: TAA at 1435
	SEQ ID NO: 10	476 aa	MW at 55285.2kD
NOV3a, CG58518-01 Protein Sequence	MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTWKQETRMKKDDSTKARP QKYEQLLHIEDNDFAMRPGFGGEYYPLKIGSPVPVGIDVHVESIDSISETNMVSFFMG YDFTMTFYLRHYWKDERLSFPSTANKSMTFDHRLTRKIWVPDIFFVHSKRSFIHDTTM ENIMLRVHPDGNVLLSLRRITVSAMCFMDFSRFPLDTQNCSLELESAYNEDDLMLYWK HGNKSLNTEEHMSLSQFFIEDFSASSGLAFYSSTGWYNRLFINFVLRRHVFFPVLQTY

104

FPAILMVMLSWVSFWIDRRAVPARVSLGGITTVLTMSTIITAVSASMPQVSYLKAVDV YLWVSSLFVFLSVIEYAAVNYLTTVEERKQFKKTGKVQISRMYNIDAVQAMAFDGCYH DSEIDMDQTSLSLNSEDFMRRKSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTY SRILFPIVYIFI

Further analysis of the NOV3a protein yielded the following properties shown in Table 3B.

Table 3B. Protein Sequence Properties NOV3a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.2400 probability located in nucleus
SignalP analysis:	Likely cleavage site between residues 25 and 26

A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 3C.

Table 3C. Geneseq Results for NOV3a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU04467	Human gamma-amino butyric acid (GABA) receptor protein #1 - Homo sapiens, 467 aa. [WO200153489-A1, 26-JUL-2001]	1..474 1..456	454/475 (95%) 454/475 (95%)	0.0
AAU04470	Human gamma-amino butyric acid (GABA) receptor protein #4 - Homo sapiens, 420 aa. [WO200153489-A1, 26-JUL-2001]	48..474 1..409	408/428 (95%) 408/428 (95%)	0.0
AAU04468	Human gamma-amino butyric acid (GABA) receptor protein #2 - Homo sapiens, 392 aa. [WO200153489-A1, 26-JUL-2001]	1..393 1..377	370/394 (93%) 370/394 (93%)	0.0
AAU04471	Human gamma-amino butyric acid (GABA) receptor protein #5 -	48..393 1..330	324/347 (93%)	e-180

105

	[WO200153489-A1, 26-JUL-2001]		(93%)	
AAU04469	Human gamma-amino butyric acid (GABA) receptor protein #3 - Homo sapiens, 180 aa. [WO200153489-A1, 26-JUL-2001]	1..192 1..177	176/192 (91%) 176/192 (91%)	2e-96

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3D.

Table 3D. Public BLASTP Results for NOV3a				
Protein Accession Number	Protein/Organism/Length	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P50573	Gamma-aminobutyric-acid receptor rho-3 subunit precursor (GABA(A) receptor) - Rattus norvegicus (Rat), 464 aa.	1..474 1..453	383/476 (80%) 407/476 (85%)	0.0
Q9YGQ2	GAMMA-AMINOBUTYRIC-ACID RECEPTOR RHO-3 SUBUNIT - Morone americana (White perch), 470 aa.	1..474 4..459	293/485 (60%) 363/485 (74%)	e-153
P50572	Gamma-aminobutyric-acid receptor rho-1 subunit precursor (GABA(A) receptor) - Rattus norvegicus (Rat), 474 aa.	49..474 58..463	270/427 (63%) 317/427 (74%)	e-144
P56475	Gamma-aminobutyric-acid receptor rho-1 subunit precursor (GABA(A) receptor) - Mus musculus (Mouse), 474 aa.	49..474 58..463	270/427 (63%) 317/427 (74%)	e-143
P24046	Gamma-aminobutyric-acid receptor rho-1 subunit precursor (GABA(A) receptor) - Homo sapiens (Human), 473 aa.	49..474 57..462	268/427 (62%) 317/427 (73%)	e-143

PFam analysis predicts that the NOV3a protein contains the domains shown in the Table 3E.

Table 3E. Domain Analysis of NOV3a			
Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Neur_chan_LBD: domain 1 of 1	88..282	70/250 (28%) 165/250 (66%)	1.2e-54
Neur_chan_memb: domain 1 of 1	289..475	44/292 (15%) 141/292 (48%)	7.6e-28

Example 4.

The NOV4 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis			
	SEQ ID NO: 11	1587 bp	
NOV4a, CG58516-01 DNA Sequence	GAACAGAAATGAATAAAAGTCGCTGGCAGAGTAGAAGACGACATGGGAGAAGAAGCCA CCAGCAGAACCCTTGGTTCAGACTCCGTGATTCTGAAGACAGGTCTGACTCCCGGGCA GCACAGCCCGCTCACGATTCCGGCCACGGTGATGACGAGTCTCCGTCAACCTCGTCTG GCACAGCTGGGACCTCCTCGTGCCAGAGCTACCTGGGTTTTACTTTGACCCTGAAAA GAAACGCTACTTCCGCTTGCTCCCTGGACATAACAACTGCAACCCCCTGACGAAAGAG AGCATCCGGCAGAAGGAGATGGAGAGCAAGAGACTGCGGCTGCTCCAGGAAGAAGACA GACGGAAAAAGATTGCCAGGATGGGATTTAATGCATCTTCCATGCTACGAAAAAGCCA GCTGGGTTTTCTCAACGTCACCAATTACTGCCATTTAGCCCACGAGCTGCGTCTCAGC TGCATGGAGAGGAAAAAGGTCCAGATTCGAAGCATGGATCCCTCCGCCTTGGCAAGCG ACCGATTTAACCTCATACTGGCAGATACCAACAGTGACCGGCTCTTCACAGTGAACGA TGTTACAGTTGGAGGCTCCAAGTATGGTATCATCAACCTGCAAAGTCTGAAGACCCCT ACGCTCAAGGTGTTCATGCCACGAAAACCTCCGATTCTCACCAACCGGAAGGTGAACA CTTCGGTGTGCTGGGCCTCGCTGAATCACTTGGATTCCCACATTCTGCTATGCCTCAT GGGACTCGCAGAGACTCCAGGCTGTGCCACCCTGCTCCCAGCATCACTGTTCGTCAAT AGTCCCCACCCAGGAATAGACCGGCCTGGCATGCTCTGCAGTTTCCGGATCCCTGGGG GTGCCTGGTCCTGTGCCTGGTCCCTGAATATCCAAGCAAATAACTGCTTCAGTACAGG CTTGTCTCGGCGGGTCCTGTTGACCAACGTGGTGACGGGACACCGGCAGTCCTTTGGG ACCAACAGTGATGTCTTGGCCCAGCAGTTTGCTCTCATGGCTCCTCTGCTGTTTAATG GCTGCCGCTCTGGGGAAATCTTTGCCATTGATCTGCGTTGTGGAAATCAAGGCAAGGG ATGGAAGGCCACCCGCCTGTTTCATGATTCAGCAGTGACCTCTGTGCGGATCCTCCAA GATGAGCAATACCTGATGGCTTCAGACATGGCTGGAAAGATCAAGCTGTGGGACCTGA GGACCACGAAGTGCGTAAGGCAGTACGAAGGCCACGTGAATGAGTACGCCTACCTGCC CCTGCATGTGCACGAGGAAGAAGGAATCCTGGTGGCAGTGGGCCAGGACTGCTACACG AGAATCTGGAGCCTCCACGATGCCCGCCTACTGAGAACCATACCCTCCCCGTACCCTG CCTCCAAGGCCGACATTCCCAGTGTGGCCTTCTCGTCGCGGCTGGGGGGGCTCCCGGGG GCGCGCCGGGGCTGCTCATGGCTGTCGGGCAGGACCTTTACTGTTACTCCTACAGCTA ATTCTGCAGGGCACAGCCCAGAGCCATGTGGATTTGACTTACGGGAGTAAAGCGTAAC TTTTTACTGCATCTAATGAGG		
	ORF Start: ATG at 9	ORF Stop: TAA at 1563	
	SEQ ID NO: 12	518 aa	MW at 57769.3kD
NOV4a, CG58516-01 Protein Sequence	MNKSRWQSRRRHGRRSHQQNPWFRLRDSEDRSDSRAAQPAHDSGHGDDESPSTSSGTA GTSSVPELPGFYFDPEKKRYFRLLPGHNNCNPLTKESIRQKEMESKRLRLLQEEDRRK KIARMGFNASSMLRKSQLGFLNVTNYCHLAHELRLSCMERKKVQIRSMDPSALASDRF NLILADTNSDRLFTVNDVTVGGSKYGIINLQSLKTPTLKVFMPRKPPILTNRKVNTSV CWASLNHLDSHILLCLMGLAETPGCATLLPASLFVNSPHPGIDRPGMLCSFRIPGGAW SCAWSLNIQANNCFSTGLSRRVLLTNVVTGHRQSFGTNSDVLAQQFALMAPLLFNGCR SGEIFAIDLRCGNQGKGWKATRLFHDSAVTSVRILQDEQYLMASDMAGKIKLWDLRTT KCVRQYEGHVNEYAYLPLHVHEEEGILVAVGQDCYTRIWSLHDARLLRTIPSPYPASK ADIPSVAFSSRLGGSRGRAGAAHGCRAGPLLLLLQLILQGTAQSHVDLTYGSKA		

5   Further analysis of the NOV4a protein yielded the following properties shown in Table 4B.

Table 4B. Protein Sequence Properties NOV4a	
Psort analysis:	0.9600 probability located in nucleus; 0.4776 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1837 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV4a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 4C.

Table 4C. Geneseq Results for NOV4a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB11794	Human secreted protein homologue, SEQ ID NO:2164 - Homo sapiens, 500 aa. [WO200157188-A2, 09-AUG-2001]	1..484 5..485	470/484 (97%) 471/484 (97%)	0.0
AAM79804	Human protein SEQ ID NO 3450 - Homo sapiens, 500 aa. [WO200157190-A2, 09-AUG-2001]	1..484 5..485	470/484 (97%) 471/484 (97%)	0.0
AAM41122	Human polypeptide SEQ ID NO 6053 - Homo sapiens, 500 aa. [WO200153312-A1, 26-JUL-2001]	1..484 5..485	470/484 (97%) 471/484 (97%)	0.0
AAG67256	Amino acid sequence of a human liver-associated gene - Homo sapiens, 489 aa. [WO200109318-A1, 08-FEB-2001]	1..484 1..474	459/484 (94%) 462/484 (94%)	0.0
AAB94587	Human protein sequence SEQ ID NO:15389 - Homo sapiens, 489 aa. [EP1074617-A2, 07-FEB-2001]	1..484 1..474	459/484 (94%) 462/484 (94%)	0.0

In a BLAST search of public sequence databases, the NOV4a protein was found to

5   have homology to the proteins shown in the BLASTP data in Table 4D.

Table 4D. Public BLASTP Results for NOV4a				
Protein Accession Number	**Protein/Organism/Length**	**NOV4a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAH18979	HYPOTHETICAL 55.7 KDA PROTEIN - Homo sapiens (Human), 495 aa.	1..484 1..480	470/484 (97%) 471/484 (97%)	0.0
Q96K22	CDNA FLJ14839 FIS, CLONE OVARC1001791 - Homo sapiens (Human), 489 aa.	1..484 1..474	459/484 (94%) 462/484 (94%)	0.0
Q9Y4P5	HYPOTHETICAL 48.5 KDA PROTEIN - Homo sapiens (Human), 430 aa (fragment).	5..435 2..428	420/431 (97%) 421/431 (97%)	0.0
Q99LF7	HYPOTHETICAL 58.1 KDA PROTEIN - Mus musculus (Mouse), 519 aa.	1..484 1..481	378/485 (77%) 423/485 (86%)	0.0
Q9UFI0	HYPOTHETICAL 26.0 KDA PROTEIN - Homo sapiens (Human), 234 aa (fragment).	269..483 4..217	175/215 (81%) 193/215 (89%)	4e-99

PFam analysis predicts that the NOV4a protein contains the domains shown in the Table 4E.

Table 4E. Domain Analysis of NOV4a			
Pfam Domain	**NOV4a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
WD40: domain 1 of 3	281..316	2/37 (5%) 26/37 (70%)	5.8e+02
WD40: domain 2 of 3	367..402	10/37 (27%) 27/37 (73%)	6.1
WD40: domain 3 of 3	408..446	10/39 (26%) 23/39 (59%)	13

Example 5.

The NOV5 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 5A.

<table>
<tr><td colspan="3">Table 5A. NOV5 Sequence Analysis</td></tr>
</table>

	SEQ ID NO: 13	1081 bp
NOV5a, CG58473-01 DNA Sequence	AGGATGGCCCAGAAGGAGAACAGTTATCCCTGGCCCTATGGCAAGCAGACGGCTCCAG CCGGCCTGAGTACCCTGCTCCCGCGAGTCCTCCCGAGGATCCCCACCGAAGCTGCGCG TGAGCTCCCGAGCTGCGCAGACCCACAGCCCGCAGCGGCCCCTGGCCATGAGGTGGTA GAGAACAGTTGTGGGAAGCGCAGCATCTTAACGCGGCCCTTCCTGGTCGACGACCTTG AGACTGGGCGTCCCCTGGGCAAAGACAAGTTTGTACATGTGTACTTGGCTCGAAAGAA GACAAGCCATTTCATCGTGGCCCTCAAGGCCTTCAAGTCTCAGATAGAGGAGGGCGTG GAGCACCAGATGCGCAGGCAGATGGAAATCCAGGCCCCCTTTCAGCATCCCAACATAT TGAGTCTCTACAACTATTTTTATGACCTGAGAAAAATCTACTGGATTCTAGAGTACGC CCCCGCCACCCCTACCCCCGAGGAGCTGTACCAGGAGCTGCGAAAGAGCCGCACCTTT GACAAGAAGCCAACAGCCACCATCACGGGGGAGGTGGCAGATGCTCTGATGTACTGCC ACGGGAAGAAGGTGACTCCCAGAGACATGAAGCCAGATAATCTACTCTCAGGGCTTGA GGGCGAGCTGAAAGTTGCCGACTTCGGCTGCCCTGTGCACGCCCCCTCACTGAGGAGG AAGACAAGACAAATGTGTGGCACCCTGGACTACCTGTCCCCAGAGACAATTGAGGGGC GCGCGCACACCGAGAAGGTGGATTTGTGGTACATCGGAGCACTCGGCTATGAGCCGCT GGTGGGGAACCCCACACACAATGAGGCCTATGGGCGAATCGTCAAGGTGGCCCTAAAA TTCCCCCTTCTGTGCCCAGGAGAGCCCCAGGACCTCATCTCCAAGCTGCTTAGGCATA ACCCCTCAGAACGGCTGCCCCTGGCCCAGGTCTCAGCCCACCCTGGGATCCTGGCCCA TTCTCGGAGGGTTTTGCCTCCCTCTGCCCATCAGTCTGTCCCCTGGTGGTCCCTGACA TTCACTCGGGGGCGTCTGTGTTTGTAAGTCTGCATAT
	ORF Start: ATG at 4	ORF Stop: TAA at 1069
	SEQ ID NO: 14	355 aa	MW at 40012.7kD
NOV5a, CG58473-01 Protein Sequence	MAQKENSYPWPYGKQTAPAGLSTLLPRVLPRIPTEAARELPSCADPQPAAAPGHEVVE NSCGKRSILTRPFLVDDLETGRPLGKDKFVHVYLARKKTSHFIVALKAFKSQIEEGVE HQMRRQMEIQAPFQHPNILSLYNYFYDLRKIYWILEYAPATPTPEELYQELRKSRTFD KKPTATITGEVADALMYCHGKKVTPRDMKPDNLLSGLEGELKVADFGCPVHAPSLRRK TRQMCGTLDYLSPETIEGRAHTEKVDLWYIGALGYEPLVGNPTHNEAYGRIVKVALKF PLLCPGEPQDLISKLLRHNPSERLPLAQVSAHPGILAHSRRVLPPSAHQSVPWWSLTF TRGRLCL

Further analysis of the NOV5a protein yielded the following properties shown in Table 5B.

<table>
<tr><td colspan="2">Table 5B. Protein Sequence Properties NOV5a</td></tr>
<tr><td>Psort analysis:</td><td>0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1897 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space</td></tr>
<tr><td>SignalP analysis:</td><td>No Known Signal Sequence Predicted</td></tr>
</table>

A search of the NOV5a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 5C.

<table>
<tr><td colspan="5">Table 5C. Geneseq Results for NOV5a</td></tr>
<tr>
<td>Geneseq Identifier</td>
<td>Protein/Organism/Length [Patent #, Date]</td>
<td>NOV5a Residues/ Match Residues</td>
<td>Identities/ Similarities for the Matched Region</td>
<td>Expect Value</td>
</tr>
<tr>
<td>AAG67615</td>
<td>Amino acid sequence of a human</td>
<td>1..341
1..343</td>
<td>247/349 (70%)
274/349 (77%)</td>
<td>e-129</td>
</tr>
</table>

110

	[WO200109316-A1, 08-FEB-2001]			
AAG67436	Amino acid sequence of a human polypeptide - Homo sapiens, 344 aa. [WO200109345-A1, 08-FEB-2001]	1..341 1..343	247/349 (70%) 274/349 (77%)	e-129
AAY22475	Human AUR1 protein sequence - Homo sapiens, 344 aa. [WO9937788-A2, 29-JUL-1999]	1..341 1..343	247/349 (70%) 274/349 (77%)	e-129
AAW18083	Human Aurora-1 - Homo sapiens, 344 aa. [WO9722702-A1, 26-JUN-1997]	1..341 1..343	247/349 (70%) 274/349 (77%)	e-129
AAY27052	Human protein kinase (HPKM)-1 (clone ID 2940) - Homo sapiens, 347 aa. [WO9938981-A2, 05-AUG-1999]	1..341 1..346	246/352 (69%) 274/352 (76%)	e-127

In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5D.

Table 5D. Public BLASTP Results for NOV5a				
Protein Accession Number	Protein/Organism/Length	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O60446	AURORA-RELATED KINASE 2 (SERINE/THREONINE KINASE 12) - Homo sapiens (Human), 344 aa.	1..341 1..343	247/349 (70%) 274/349 (77%)	e-128
Q96GD4	UNKNOWN (PROTEIN FOR MGC:11031) - Homo sapiens (Human), 344 aa.	1..341 1..343	247/349 (70%) 274/349 (77%)	e-128
Q96DV5	UNKNOWN (PROTEIN FOR MGC:4243) - Homo sapiens (Human), 345 aa.	1..341 1..344	247/350 (70%) 274/350 (77%)	e-126
Q9UQ46	AIK2 – Homo sapiens (Human), 343 aa.	1..341 1..342	245/348 (70%) 272/348 (77%)	e-126
O14630	PROTEIN KINASE - Homo sapiens (Human), 347 aa.	1..341 1..346	245/352 (69%) 272/352 (76%)	e-125

111

PFam analysis predicts that the NOV5a protein contains the domains shown in the Table 5E.

Table 5E. Domain Analysis of NOV5a			
Pfam Domain	**NOV5a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Pkinase: domain 1 of 1	76..325	81/293 (28%) 184/293 (63%)	6.5e-36

Example 6.

The NOV6 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis		
	SEQ ID NO: 15	1524 bp
NOV6a, CG58470-01 DNA Sequence	AGCATTATGAACACTAATGACCTTAAACTCAGGTTGTCCAAAGCTGAGCAAGAACACC CACTACGTTTCTGGAATGAGCTTGAAGAAGCCCGACAGGTAGAACTTTATGCAGAGCT CCAGGCCATCGACTTTCAGGAACTGAACTTCTTTTTCCAAAAGGCCATTGAAGGATTT AACCAGTCCTCTCATCAAGAAAAGGTGGATGCGGGAATGGAACCTGTCCCTCGAGAAG TACTGGGCAGTGCTGCAGGGAAGCTAGATCAGCTCCAGGCCTGGGAAAGCAAAGTTTT CCAGATTTCTGAGAACAAAGTCACAGTTGTTCTAGCTGGTGGGCAGGGGACTAGACTC GTTGCATATCCAAAGGGGATGTATGATGTTGGTTTGCCATCCCATAAGACACTTTTTC AGATTCAAGCAGAGCATATCCTGAAGCTACAACAGTTAGCTGAAAAATATTATGGCAA CAAATGCATTATTCCATATTACGTCATGACCAGCGAGTTCACTCTGGGGCCCACGGCC GAGTTCTTCAGGGAGCACAACTTCTTCCACCTGGACCCCGCCAACGTGGTCATGTTTG AGCAGCGCCTGCTGCCTGCTGTGACCTTTGATGGCAAGGTTATCCTGGAGCGGAAAGA CAAAGTTGCCATGGCCCCAGACGGCAACGGGGGCCTCTACTGCGCGCTGGAGGACCAC AAGATCCTGGAGGACATGGAGCGCCGGGGAGTGGAGTTTGTGCACGTGTACTGTGTGG ACAACATCCTGGTGCGGCTGGCGGACCCTGTCTTCATCGGCTTCTGTGTGTTGCAGGG CGCAGACTGTGGCGCCAAGGTGGTGGAAAAGGCATACCCCGAGGAGCCCGTGGGCGTG GTGTGCCAGGTGGACGGTGTCCCCCAGGTGGTGGAGTACAGCGAGATCAGTCCTGAGA CCGCACAGCTACGTGTCTCCGACGGGAGCCTGCTGTACAATGCAGGCAACATCTGCAA CCACTTCTTCACCCGAGGCTTCCTTAAGGCGGTCACCAGGGAGTTTGAGCCTTTGCTG AAGCCACACGTGGCTGTGAAGAAGGTCCCGTATGTGGATGAGGAGGGGAATCTGGTAA AGCCGCTAAAACCGAACGGGATAAAGATGGAGAAGTTTGTGTTTGATGTGTTCCGGTT TGCTAAGAACTTTGCTGCCTTGGAAGTGCTGCGGGAGGAGGAATTTTCCCCACTGAAG AACGCAGAGCCAGCCGACAGGGACAGTCCCCGCCACCGCTCGCCAGGCCCTGCTCACCC AGCACTACCGGTGGGCTCTGCGGGCCGGGGCCCGCTTCCTGGATGCCCATGGGGCCCG GCTCCCAGAGCTGCCCAGCTTGCCCCCAAATGGAGACCCTCCGGCCATCTGTGAGATA TCGCCCTTGGTGTCTTACTCTGGAGAGGGTTTAGAAGTGTACCTGCAAGGCCGGGAGT TCCAGTCCCCGCTCATCCTGGATGAAGACCAGGCCAGGGAGCTGGTGAAAAATGGTAT ATGAACCTGATACCAA
	ORF Start: ATG at 7	ORF Stop: TGA at 1510
	SEQ ID NO: 16	501 aa · MW at 56461.0kD
NOV6a, CG58470-01 Protein Sequence	MNTNDLKLRLSKAEQEHPLRFWNELEEARQVELYAELQAIDFQELNFFFQKAIEGFNQ SSHQEKVDAGMEPVPREVLGSAAGKLDQLQAWESKVFQISENKVTVVLAGGQGTRLVA YPKGMYDVGLPSHKTLFQIQAEHILKLQQLAEKYYGNKCIIPYYVMTSEFTLGPTAEF FREHNFFHLDPANVVMFEQRLLPAVTFDGKVILERKDKVAMAPDGNGGLYCALEDHKI LEDMERRGVEFVHVYCVDNILVRLADPVFIGFCVLQGADCGAKVVEKAYPEEPVGVVC QVDGVPQVVEYSEISPETAQLRVSDGSLLYNAGNICNHFFTRGFLKAVTREFEPLLKP HVAVKKVPYVDEEGNLVKPLKPNGIKMEKFVFDVFRFAKNFAALEVLREEEFSPLKNA EPADRDSPRTARQALLTQHYRWALRAGARFLDAHGARLPELPSLPPNGDPPAICEISP LVSYSGEGLEVYLQGREFQSPLILDEDQARELVKNGI

Further analysis of the NOV6a protein yielded the following properties shown in Table 6B.

Table 6B. Protein Sequence Properties NOV6a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3490 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 6C.

Table 6C. Geneseq Results for NOV6a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB56960	Human prostate cancer antigen protein sequence SEQ ID NO:1538 - Homo sapiens, 524 aa. [WO200055174-A1, 21-SEP-2000]	1..501 3..524	353/522 (67%) 413/522 (78%)	0.0
AAG32392	Arabidopsis thaliana protein fragment SEQ ID NO: 39067 - Arabidopsis thaliana, 502 aa. [EP1033405-A2, 06-SEP-2000]	9..485 36..497	194/489 (39%) 275/489 (55%)	3e-84
AAG40236	Arabidopsis thaliana protein fragment SEQ ID NO: 49896 - Arabidopsis thaliana, 477 aa. [EP1033405-A2, 06-SEP-2000]	9..485 12..472	193/488 (39%) 272/488 (55%)	3e-82
AAG40235	Arabidopsis thaliana protein fragment SEQ ID NO: 49895 - Arabidopsis thaliana, 500 aa. [EP1033405-A2, 06-SEP-2000]	9..485 35..495	193/488 (39%) 272/488 (55%)	3e-82
AAG40234	Arabidopsis thaliana protein fragment SEQ ID NO: 49894 - Arabidopsis thaliana, 505 aa. [EP1033405-A2, 06-SEP-2000]	9..485 40..500	193/488 (39%) 272/488 (55%)	3e-82

5   In a BLAST search of public sequence databases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table 6D.

Table 6D. Public BLASTP Results for NOV6a				
Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96GM2	UDP-N-ACTEYLGLUCOSAMINE PYROPHOSPHORYLASE 1 - Homo sapiens (Human), 505 aa.	1..501 1..505	351/505 (69%) 412/505 (81%)	0.0
Q16222	UDP-N-acetylhexosamine pyrophosphorylase (Antigen X) (AGX) (Sperm- associated antigen 2) [Includes: UDP-N-acetylgalactosamine pyrophosphorylase (EC 2.7.7.-) (AGX-1); UDP-N-acetylglucosamine pyrophosphorylase (EC 2.7.7.23) (AGX-2)] - Homo sapiens (Human), 522 aa.	1..501 1..522	352/522 (67%) 412/522 (78%)	0.0
Q91YN5	HYPOTHETICAL 58.6 KDA PROTEIN - Mus musculus (Mouse), 522 aa.	1..501 1..522	342/522 (65%) 407/522 (77%)	0.0
AAH17547	HYPOTHETICAL 58.5 KDA PROTEIN - Mus musculus (Mouse), 521 aa.	1..501 1..521	341/521 (65%) 407/521 (77%)	0.0
Q9Y0Z0	BCDNA:LD24639 PROTEIN - Drosophila melanogaster (Fruit fly), 520 aa.	6..492 44..513	236/491 (48%) 330/491 (67%)	e-124

PFam analysis predicts that the NOV6a protein contains the domains shown in the Table 6E.

Table 6E. Domain Analysis of NOV6a			
Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
UDPGP: domain 1 of 1	40..434	108/428 (25%) 324/428 (76%)	8.4e-111

Example 7.

5   The NOV7 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 7A.

Table 7A. NOV7 Sequence Analysis		
	SEQ ID NO: 17	461 bp
NOV7a, CG58593-01 DNA Sequence	ACGCAGAGATGCAGATCTTTGTGAAGACCCTCACGGGCAAGACCATCACCCTTGAGGT CAAGCCCACCGACACCATTGAGAATGTCAAAACCAAAATTCAGGACAAGGAGGGTATC CCACCTGACCAGCAGCGTCTGATATTTGCTGGGAAACGGCTGGAGGATGGCCACACTC TCTCAGGCTACAACATCCAGAAAGAGTCCACCCTAAACCTGGTGCTGCGCCTGCGAGG TGGCATTACTGAGCCTTCCCTCCGCCAGCTCGTCCAGAAATACAACTGCGACGAGATG ATCTGCTGCAAGTGCTATGCTTGCCTGCACCCCGGTGCTATCAACTGCCACAAGAAGA AATGCGGCCACACCAACAACCTGTACCCCAGGAAGAAGGTCAAATAAGGCTCTTCCTT CCTTGAAGGGCAGCAGCCTTCTGCCCAGGCCCCATGGCCCTGGGGCCTCAATAAA	
	ORF Start: ATG at 9	ORF Stop: TAA at 393
	SEQ ID NO: 18	128 aa	MW at 14540.9kD
NOV7a, CG58593-01 Protein Sequence	MQIFVKTLTGKTITLEVKPTDTIENVKTKIQDKEGIPPDQQRLIFAGKRLEDGHTLSG YNIQKESTLNLVLRLRGGITEPSLRQLVQKYNCDEMICCKCYACLHPGAINCHKKKCG HTNNLYPRKKVK		

Further analysis of the NOV7a protein yielded the following properties shown in Table 7B.

Table 7B. Protein Sequence Properties NOV7a	
PSort analysis:	0.9800 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV7a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 7C.

Table 7C. Geneseq Results for NOV7a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB52080	Gene 16 human secreted protein homologous amino acid sequence #129 - Sus scrofa, 128 aa. [WO200061596-A1, 19-OCT-2000]	1..128 1..128	111/128 (86%) 118/128 (91%)	7e-61
AAG43861	Arabidopsis thaliana protein fragment SEQ ID NO: 54871 - Arabidopsis thaliana, 128 aa. [EP1033405-A2, 06-SEP-2000]	1..128 1..128	101/128 (78%) 113/128 (87%)	9e-55

115

AAG36188	Arabidopsis thaliana protein fragment SEQ ID NO: 44314 - Arabidopsis thaliana, 249 aa. [EP1033405-A2, 06-SEP-2000]	1..128 122..249	101/128 (78%) 113/128 (87%)	9e-55
AAG36187	Arabidopsis thaliana protein fragment SEQ ID NO: 44313 - Arabidopsis thaliana, 264 aa. [EP1033405-A2, 06-SEP-2000]	1..128 137..264	101/128 (78%) 113/128 (87%)	9e-55
AAG36186	Arabidopsis thaliana protein fragment SEQ ID NO: 44312 - Arabidopsis thaliana, 322 aa. [EP1033405-A2, 06-SEP-2000]	1..128 195..322	101/128 (78%) 113/128 (87%)	9e-55

In a BLAST search of public sequence databases, the NOV7a protein was found to have homology to the proteins shown in the BLASTP data in Table 7D.

Table 7D. Public BLASTP Results for NOV7a				
Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BX98	UBIQUITIN A-52 RESIDUE RIBOSOMAL PROTEIN FUSION PRODUCT 1 - Homo sapiens (Human), 141 aa (fragment).	1..128 14..141	111/128 (86%) 118/128 (91%)	3e-60
Q9UPK7	UBIQUITIN-52 AMINO ACID FUSION PROTEIN - Homo sapiens (Human), 128 aa.	1..128 1..128	111/128 (86%) 118/128 (91%)	3e-60
Q9PT09	UBIQUITIN - Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri), 128 aa.	1..128 1..128	110/128 (85%) 118/128 (91%)	6e-60
O42388	UBIQUITIN-RIBOSOMAL PROTEIN FUSION PROTEIN - Gallus gallus (Chicken), 128 aa.	1..128 1..128	110/128 (85%) 117/128 (90%)	7e-60
Q9XSV1	UBIQUITIN-RIBOSOMAL PROTEIN L40 FUSION PROTEIN - Canis familiaris (Dog), 128 aa.	1..128 1..128	110/128 (85%) 117/128 (90%)	1e-59

PFam analysis predicts that the NOV7a protein contains the domains shown in the

5   Table 7E.

116

Table 7E. Domain Analysis of NOV7a

Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ubiquitin: domain 1 of 1	1..74	54/83 (65%) 72/83 (87%)	1.9e-38
Ribosomal_L40e: domain 1 of 1	77..128	30/52 (58%) 42/52 (81%)	7.3e-20

Example 8.

The NOV8 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 8A.

Table 8A. NOV8 Sequence Analysis

	SEQ ID NO: 19	2296 bp	
NOV8a, CG57871-01 DNA Sequence	CGGCGGCGGCGGCAGTAGAAATGATGGAAGAATTGCATAGCCTGGACCCACGACGGCA GAAATTATTGGAGGCCAGGTTTACTGGAGTAGGTGTTAGTAAGGGACCACTTAATAGT GAGTCTTCCAACCAGAGCTTGTGCAGCGTCGGATCCTTGAGTGATAAAGAAGTAGAGA CTCCCAAGAAAAAGCAGAATGACCAGCGAAATCGGAAAAGAAAAGCTGAACCATATGA AAGTAGCCAAGGGAAAGGCACTCCTAGGGGACATAAAATTAGTGATTACTTTGAGTTT GCTGGGGGAAGCGGGCCGGGAACCAGCCCTGGCAGAAGTGTTCCACCAGTTGCACGAT CCTCACTGCAACATTCTTTATCCAATCCCTTACCGCGACGAGTAGAACAGCCCCTCTA TGGTTTAGATGGCAGTGCTGCAAAGGAGGCAACGGAGGAGCAGTCTGCTCTGCCAACC CTCATGTCAGTGATGCTAGCAAAACCTCGGCTTGACACAGAGCAGCTGGCGCAAAGGG GAGCTGGCCTCTGCTTCACTTTTGTTTCAGCTCAGCAAAACAGTCCCTCATCTACGGG ATCTGGCAACACAGAGCATTCCTGCAGCTCCCAAAAACAGATCTCCATCCAGCACAGA CAGACCCAGTCCGACCTCACAATAGAAAAAATATCTGCACTAGAAAACAGTAAGAATT CTGACTTAGAGAAGAAGGAGGGAAGAATAGATGATTTATTAAGAGCCATCTGTGATTT GAGACGGCAGATTGATGAACAGCAAAAGATGCTAGAGAAATACAAGGAACGATTAAAT AGATGTGTGACAATGAGCAAGAAACTCCTTATAGAAAAGTCAAAACAAGAGAAGATGG CGTGTAGAGATAAGAGCATGCAAGACCGCTTGAGACTGGGCCACTTTACTACGTCTGA CCACGGAGCCAAATTTACTGAGCAGTGGACAGATGGTTATGCTTTTCAGAATCTTATC AAGCAACAGGAAAGGATAAATTCACAGAGGGAAGAGATAGAAAGACAACGGAAAATGT TAGCAAAGCGGAAACCTCCTGCCATGGGTCAGGCCCCTCCTGCAACCAATGAGCAGAA ACAGTGGAAAAGCAAGACCAATGGAGCTGAAAATGAAACGTTAACGTTAAAAGAATAC CATGAACAAGAAGAAATCTTCAAACTCAGATTAGGTCATCTTAAAAAGGAGGAAGCAG AGATCCAGGCAGAGCTGGAGAGGCTAGAAAGGGTTAGAAAACTACATATCAGGGAAGT AAAAAGGATACATAATGAAGATAATTCACAATTTAAATATCATCCAACGCTAAATGAC AGATATTTGTTGTTACATCTTTTGGGTAGAGGAGGTTTCAGTGAAGTTTACAAGGCAT TTGATCTAACAGAGCAAAGATACGTAGCTGTGAAAATTCACCAGTTAAATAAAAACTG GAGAGATGAGAAAAAGGAGAATTACCACAAGCATGCATGTAGGGAATACCGGATTCAT AAAGAGCTGGACCATCCCAGAATAGTTAAGCTGTATGATTACTTTTCACTGGATACTG ACTCGTTTTGTACAGTATTAGAATACTGTGAGGGAAATGATCTGGACTTCTACCTGAA ACAGCACAAATTAATGTCAGAGAAAGAGGCCCGGTCCATTATCATGCAGATTGTGAAT GCTTTAAAGTACTTAAATGAAATAAAACCTCCCATCATACACTATGACCTCAAACCAG GTAATATTCTTTTAGAAAATGGTACAGCGTGTGGAGAGATAAAAATTACAGATTTTGG TCTTTCGAAGATCATGGATGATGATAGCTACAATTCAGTGGATGGCATGGAGCTAACA TCACAAGGTGCTGGTACTTATTGGTATTTACCACCAGAGTGTTTTGTGGTTGGGAAAG AACCACCAAAGATCTCAAATAAAGTTGATGTGTGGTCGGTGGGTGTGATCTTCTATCA GTGTCTTTATGGAAGGAAGCCTTTTGGCCATAACCAGTCTCAGCAAGACATCCTACAA GAGAATACGATTCTTAAAGCTACTGAAGTGCAGTTCCCGCCAAAGCCGGTAGTAACAC CTGAAGCAAAGGCGTTGATTCGACGATGCTTGGCCTACCGAAAGGAGGACCGCATTGA TGTCCAGCAGCTGGCCTGTGATCCCTACTTGTTGCCTCACATCCGAAAGTCAGTCTCT ACGAGTAGCCCTGCTGGAGCTGCTATTGCATCAACCTCTGGGGCGTCCAATAACAGTT CTTCTAATTGAGACTGACTCCAAGGCCACAAACT		
	ORF Start: ATG at 24	ORF Stop: TGA at 2271	
	SEQ ID NO: 20	749 aa	MW at 85415.8kD
NOV8a, CG57871-01 Protein Sequence	MEELHSLDPRRQKLLEARFTGVGVSKGPLNSESSNQSLCSVGSLSDKEVETPKKKQND QRNRKRKAEPYESSQGKGTPRGHKISDYFEFAGGSGPGTSPGRSVPPVARSSLQHSLS NPLPRRVEQPLYGLDGSAAKEATEEQSALPTLMSVMLAKPRLDTEQLAQRGAGLCFTF		

117

```
VSAQQNSPSSTGSGNTEHSCSSQKQISIQHRQTQSDLTIEKISALENSKNSDLEKKEG
RIDDLLRAICDLRRQIDEQQKMLEKYKERLNRCVTMSKKLLIEKSKQEKMACRDKSMQ
DRLRLGHFTTSDHGAKFTEQWTDGYAFQNLIKQQERINSQREEIERQRKMLAKRKPPA
MGQAPPATNEQKQWKSKTNGAENETLTLKEYHEQEEIFKLRLGHLKKEEAEIQAELER
LERVRKLHIREVKRIHNEDNSQFKYHPTLNDRYLLLHLLGRGGFSEVYKAFDLTEQRY
VAVKIHQLNKNWRDEKKENYHKHACREYRIHKELDHPRIVKLYDYFSLDTDSFCTVLE
YCEGNDLDFYLKQHKLMSEKEARSIIMQIVNALKYLNEIKPPIIHYDLKPGNILLENG
TACGEIKITDFGLSKIMDDDSYNSVDGMELTSQGAGTYWYLPPECFVVGKEPPKISNK
VDVWSVGVIFYQCLYGRKPFGHNQSQQDILQENTILKATEVQFPPKPVVTPEAKALIR
RCLAYRKEDRIDVQQLACDPYLLPHIRKSVSTSSPAGAAIASTSGASNNSSSN
```

Further analysis of the NOV8a protein yielded the following properties shown in Table 8B.

Table 8B. Protein Sequence Properties NOV8a

PSort analysis:	0.9600 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 8C.

Table 8C. Geneseq Results for NOV8a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM39278	Human polypeptide SEQ ID NO 2423 - Homo sapiens, 718 aa. [WO200153312-A1, 26-JUL-2001]	1..749 2..718	703/749 (93%) 707/749 (93%)	0.0
AAM41064	Human polypeptide SEQ ID NO 5995 - Homo sapiens, 809 aa. [WO200153312-A1, 26-JUL-2001]	1..749 92..809	695/750 (92%) 701/750 (92%)	0.0
AAR76062	Protein kinase PKU beta - Homo sapiens, 540 aa. [JP07132093-A, 23-MAY-1995]	210..749 1..540	525/540 (97%) 527/540 (97%)	0.0
AAR76061	Protein kinase PKU alpha - Homo sapiens, 787 aa. [JP07132093-A, 23-MAY-1995]	1..744 49..783	537/794 (67%) 592/794 (73%)	0.0
ABB20910	Protein #2909 encoded by probe	346..749 1..404	404/404 (100%) 404/404 (100%)	0.0

118

	expression - Homo sapiens, 404 aa. [WO200157274-A2, 09-AUG-2001]			

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8D.

Table 8D. Public BLASTP Results for NOV8a				
Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UKI7	TOUSLED-LIKE KINASE 2 - Homo sapiens (Human), 749 aa.	1..749 1..749	731/749 (97%) 736/749 (97%)	0.0
O55047	TOUSLED-LIKE KINASE - Mus musculus (Mouse), 717 aa.	1..749 1..717	699/749 (93%) 705/749 (93%)	0.0
Q9Y4F7	PKU-ALPHA - Homo sapiens (Human), 719 aa (fragment).	1..749 3..719	700/749 (93%) 705/749 (93%)	0.0
Q9D5Y5	TOUSLED-LIKE KINASE 2 (ARABIDOPSIS) - Mus musculus (Mouse), 696 aa.	1..656 1..656	629/656 (95%) 640/656 (96%)	0.0
Q90ZY7	PKU-ALPHA PROTEIN KINASE - Brachydanio rerio (Zebrafish) (Zebra danio), 697 aa.	1..749 2..697	580/753 (77%) 626/753 (83%)	0.0

PFam analysis predicts that the NOV8a protein contains the domains shown in the
5   Table 8E.

Table 8E. Domain Analysis of NOV8a			
Pfam Domain	NOV8a Match Region	Identities/ Similarities for the Matched Region	Expect Value
A2M: domain 1 of 1	501..523	10/23 (43%) 20/23 (87%)	4.6
Pkinase: domain 1 of 1	439..718	96/316 (30%) 213/316 (67%)	5.4e-70

Example 9.

The NOV9 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 9A.

<table>
<tr><td colspan="4" align="center">Table 9A. NOV9 Sequence Analysis</td></tr>
<tr><td></td><td colspan="3">SEQ ID NO: 21   2060 bp</td></tr>
<tr><td>NOV9a,
CG58590-01 DNA Sequence</td><td colspan="3">
GTTTTCATAGATAACCATGACAACATCCCATATGAATGGGCATGTTACAGAGGAATCA
GACAGCGAAGTAAAAAATGTTGATCTTGCATCACCAGAGGAACATCAGAAGCACCGAG
AGATGGCTGTTGACTGCCCTGGAGATTTGGGCACCAGGATGATGCCAATACGTCGAAG
TGCACAGTTGGAGCGTATTCGGCAACAACAGGAGGACATGAGGCGTAGGAGAGAGGAA
GAAGGGAAAAAGCAAGAACTTGACCTTAATTCTTCCATGAGACTTAAGAAACTAGCCC
AAATTCCTCCAAAGACCGGAATAGATAACCCTATGTTTGATACAGAGGAAGGAATTGT
CTTAGAAAGTCCTCATTATGCTGTGAAAATATTAGAAATAGAAGACTTGTTTTCTTCA
CTTAAACATATCCAACATACTTTGGTAGATTCTCAGAGCCAGGAGGATATTTCACTGC
TTTTACAACTTGTTCAAAATAAGGATTTCCAGAATGCATTTAAGATACACAATGCCAT
CACAGTACACATGAACAAGGCCAGTCCTCCATTTCCTCTTATCTCCAACGCACAAGAT
CTTGCTCAAGAGGTACAAACTGTTTTGAAGCCAGTTCATCATAAGGAAGGACAAGAAC
TAACTGCTTTGCTGAATACTCCACATATTCAGGCACTTTTACTGGCCCACGATAAGGT
TGCTGAGCAGGAAATGCAGCTAGAGCCCATTACAGATGAGAGAGTTTATGAAAGTATT
GGCCAGTATGGAGGAGAAACTGTAAAAATAGTTCGTATAGAAAAGGCTCGTGATATTC
CGTTGGGTGCTACAGTTCGTAATGAAATGGACTCTGTCATCATTAGCCGGATAGTAAA
AGGGGGTGCTGCAGAGAAAAGTGGTCTGTTGCATGAAGGAGATGAAGTTCTAGAGATT
AATGGCATTGAAATTCGGGGGAAAGATGTCAATGAGGTTTTTGACTTGTTGTCTGATA
TGCATGGTACTTTGACTTTTGTCCTGATTCCCAGTCAACAGATCAAGCCGCCTCCTGC
CAAGGAAACAGTAATCCATGTAAAAGCTCATTTTGACTATGACCCCTCAGATGACCCT
TATGTTCCATGTCGAGAGTTAGGTCTGTCTTTTCAAAAAGGTGATATACTTCATGTGA
TCAGTCAAGAAGATCCAAACTGGTGGCAGGCCTACAGGGAAGGGGACGAAGATAATCA
ACCTCTAGCCGGGCTTGTTCCAGGGAAAAGCTTTCAGCAGCAAAGGGAAGCCATGAAA
CAAACCATAGAAGAAGATAAGGAGCCAGAAAAATCAGGTAAACTGTGGTGTGCAAAGA
AGAATAAAAAGAAGAGGAAAAAGGTTTTATATAATGCCAATAAAAATGATGATTATGA
CAACGAGGAGATCTTAACCTATGAGGAAATGTCACTTTATCATCAGCCAGCAAATAGG
AAGAGACCTATCATCTTGATTGGTCCACAGAACTGTGGCCAGAATGAATTGCGTCAGA
GGCTCATGAACAAAGAAAAGGACCGCTTTGCATCTGCAGTTCCTCGTACAACCCGGAG
TAGGCGAGACCAAGAAGTAGCCGGTAGAGATTACCACTTTGTTTCGCGGCAAGCATTC
GAGGCAGACATAGCAGCTGGAAAGTTCATTGAGCATGGTGAATTTGAGAAGAATTTGT
ATGGAACTAGCATAGATTCTGTACGGCAAGTGATCAACTCTGGCAAAATATGTCTTTT
AAGTCTTCGTACACAGTCATTGAAGACTCTCCGGAATTCAGATTTGAAACCATATATT
ATCTTCATTGCACCCCCTTCACAAGAAAGACTTCGGGCATTATTGGCCAAAGAAGGCA
AGAATCCAAAGCCTGAAGAGTTGAGAGAAATCATTGAGAAGACAAGAGAGATGGAGCA
GAACAATGGCCACTACTTTGATACGGCAATTGTGAATTCCGATCTTGATAAAGCCTAT
CAGGAATTGCTTAGGTTAATTAACAAACTTGATACTGAACCTCAGTGGGTACCATCCA
CTTGGCTGAGGTGAAAGAAACATCCATTCT
</td></tr>
<tr><td></td><td colspan="3">ORF Start: ATG at 17   ORF Stop: TGA at 2042</td></tr>
<tr><td></td><td>SEQ ID NO: 22</td><td>675 aa</td><td>MW at 77311.8kD</td></tr>
<tr><td>NOV9a,
CG58590-01 Protein Sequence</td><td colspan="3">
MTTSHMNGHVTEESDSEVKNVDLASPEEHQKHREMAVDCPGDLGTRMMPIRRSAQLER
IRQQQEDMRRRREEEGKKQELDLNSSMRLKKLAQIPPKTGIDNPMFDTEEGIVLESPH
YAVKILEIEDLFSSLKHIQHTLVDSQSQEDISLLLQLVQNKDFQNAFKIHNAITVHMN
KASPPFPLISNAQDLAQEVQTVLKPVHHKEGQELTALLNTPHIQALLLAHDKVAEQEM
QLEPITDERVYESIGQYGGETVKIVRIEKARDIPLGATVRNEMDSVIISRIVKGGAAE
KSGLLHEGDEVLEINGIEIRGKDVNEVFDLLSDMHGTLTFVLIPSQQIKPPPAKETVI
HVKAHFDYDPSDDPYVPCRELGLSFQKGDILHVISQEDPNWWQAYREGDEDNQPLAGL
VPGKSFQQQREAMKQTIEEDKEPEKSGKLWCAKKNKKKRKKVLYNANKNDDYDNEEIL
TYEEMSLYHQPANRKRPIILIGPQNCGQNELRQRLMNKEKDRFASAVPRTTRSRRDQE
VAGRDYHFVSRQAFEADIAAGKFIEHGEFEKNLYGTSIDSVRQVINSGKICLLSLRTQ
SLKTLRNSDLKPYIIFIAPPSQERLRALLAKEGKNPKPEELREIIEKTREMEQNNGHY
FDTAIVNSDLDKAYQELLRLINKLDTEPQWVPSTWLR
</td></tr>
<tr><td></td><td colspan="3">SEQ ID NO: 23   2030 bp</td></tr>
<tr><td>NOV9b,
CG58590-02 DNA Sequence</td><td colspan="3">
CCATGACAACATCCCATATGAATGGGCATGTTACAGAGGAATCAGACAGCGAAGTAAA
AAATGTTGATCTTGCATCACCAGAGGAACATCAGAAGCACCGAGAGATGGCTGTTGAC
TGCCCTGGAGATTTGGGCACCAGGATGATGCCAATACGTCGAAGTGCACAGTTGGAGC
GTATTCGGCAACAACAGGAGGACATGAGGCGTAGGAGAGAGGAAGAAGGGAAAAAGCA
AGAACTTGACCTTAATTCTTCCATGAGACTTAAGAAACTAGCCCAAATTCCTCCAAAG
ACCGGAATAGATAACCCTATGTTTGATACAGAGGAAGGAATTGTCTTAGAAAGTCCTC
ATTATGCTGTGAAAATATTAGAAATAGAAGACTTGTTTTCTTCAAACATATCCA
ACATACTTTGGTAGATTCTCAGAGCCAGGAGGATATTTCACTGCTTTTACAACTTGTT
CAAAATAAGGATTTCCAGAATGCATTTAAGATACACAATGCCATCACAGTACATATGA
</td></tr>
</table>

120

ACAAGGCCAGTCCTCCATTTCCTCTTATCTCCAACGCACAAGATCTTGCTCAAGAGGT ACAAACTGTTTTGAAGCCAGTTCATCATAAGGAAGGACAAGAACTAACTGCTTTGCTG AATACTCCACATATTCAGGCACTTTTACTGGCCCACGATAAGGTTGCTGAGCAGGAAA TGCAGCTAGAGCCCATTACAGATGAGAGAGTTTATGAAAGTATTGGCCAGTATGGAGG AGAAACTGTAAAAATAGTTCGTATAGAAAAGGCTCGTGATATTCCGTTGGGTGCTACA GTTCGTAATGAAATGGACTCTGTCATCATTAGCCGGATAGTAAAAGGGGGTGCTGCAG AGAAAAGTGGTCTGTTGCATGAAGGAGATGAAGTTCTAGAGATTAATGGCATTGAAAT TCGGGGGAAAGATGTCAATGAGGTTTTTGACCTGTTGTCTGATATGCATGGTACTTTG ACTTTTGTCCTGATTCCCAGTCAACAGATCAAGCCGCCTCCTGCCAAGGAAACAGTAA TCCATGTAAAAGCTCATTTTGACTATGACCCCTCAGATGACCCTTATGTTCCATGTCG AGAGTTAGGTCTGTCTTTTCAAAAAGGTGATATACTTCATGTGATCAGTCAAGAAGAT CCAAACTGGTGGCAGGCCTACAGGGAAGGGGACGAAGATAATCAACCTCTAGCCGGGC TTGTTCCAGGGAAAAGCTTTCAGCAGCAAAGGGAAGCCATGAAACAAACCATAGAAGA AGATAAGGAGCCAGAAAAATCAGGAAAACTGTGGTGTGCAAAGAAGAATAAAAAGAAG AGGAAAAAGGTTTTATATAATGCCAATAAAAATGATGATTATGACAACGAGGAGATCT TAACCTATGAGGAAATGTCACTTTATCATCAGCCAGCAAATAGGAAGAGACCTATCAT CTTGATTGGTCCACAGAACTGTGGCCAGAATGAATTGCGTCAGAGGCTCATGAACAAA GAAAAGGACCGCTTTGCATCTGCAGTTCCTCATACAACCCGGAGTAGGCGAGACCAAG AAGTAGCCGGTAGAGATTACCACTTTGTTTCGCGGCAAGCATTCGAGGCAGACATAGC AGCTGGAAAGTTCATTGAGCATGGTGAATTTGAGAAGAATTTGTATGGAACTAGCATA GATTCTGTACGGCAAGTGATCAACTCTGGCAAAATATGTCTTTTAAGTCTTCGTACAC AGTCATTGAAGACTCTCCGGAATTCAGATTTGAAACCATATATTATCTTCATTGCACC CCCTTCACAAGAAAGACTTCGGGCATTATTGGCCAAAGAAGGCAAGAATCCAAAGCCT GAAGAGTTGAGAGAAATCATTGAGAAGACAAGAGAGATGGAGCAGAACAATGGCCACT ACTTTGATACGGCAATTGTGAATTCCGATCTTGATAAAGCCTATCAGGAATTGCTTAG GTTAATTAACAAACTTGATACTGAACCTCAGTGGGTACCATCCACTTGGCTGAGGTGA
ORF Start: ATG at 3   ORF Stop: TGA at 2028
SEQ ID NO: 24   675 aa   MW at 77292.8kD

| NOV9b,
CG58590-02 Protein Sequence | MTTSHMNGHVTEESDSEVKNVDLASPEEHQKHREMAVDCPGDLGTRMMPIRRSAQLER
IRQQQEDMRRRREEEGKKQELDLNSSMRLKKLAQIPPKTGIDNPMFDTEEGIVLESPH
YAVKILEIEDLFSSLKHIQHTLVDSQSQEDISLLLQLVQNKDFQNAFKIHNAITVHMN
KASPPFPLISNAQDLAQEVQTVLKPVHHKEGQELTALLNTPHIQALLLAHDKVAEQEM
QLEPITDERVYESIGQYGGETVKIVRIEKARDIPLGATVRNEMDSVIISRIVKGGAAE
KSGLLHEGDEVLEINGIEIRGKDVNEVFDLLSDMHGTLTFVLIPSQQIKPPPAKETVI
HVKAHFDYDPSDDPYVPCRELGLSFQKGDILHVISQEDPNWWQAYREGDEDNQPLAGL
VPGKSFQQQREAMKQTIEEDKEPEKSGKLWCAKKNKKKRKKVLYNANKNDDYDNEEIL
TYEEMSLYHQPANRKRPIILIGPQNCGQNELRQRLMNKEKDRFASAVPHTTRSRRDQE
VAGRDYHFVSRQAFEADIAAGKFIEHGEFEKNLYGTSIDSVRQVINSGKICLLSLRTQ
SLKTLRNSDLKPYIIFIAPPSQERLRALLAKEGKNPKPEELREIIEKTREMEQNNGHY
FDTAIVNSDLDKAYQELLRLINKLDTEPQWVPSTWLR |

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 9B.

Table 9B. Comparison of NOV9a against NOV9b.		
Protein Sequence	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV9b	1..675 1..675	636/675 (94%) 636/675 (94%)

Further analysis of the NOV9a protein yielded the following properties shown in Table 9C.

Table 9C. Protein Sequence Properties NOV9a	
PSort analysis:	0.7000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)

SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 9D.

Table 9D. Geneseq Results for NOV9a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB94180	Human protein sequence SEQ ID NO:14494 - Homo sapiens, 503 aa. [EP1074617-A2, 07-FEB-2001]	173..675 1..503	501/503 (99%) 501/503 (99%)	0.0
AAB41921	Human ORFX ORF1685 polypeptide sequence SEQ ID NO:3370 - Homo sapiens, 269 aa. [WO200058473-A2, 05-OCT-2000]	406..675 1..269	261/270 (96%) 264/270 (97%)	e-147
AAU07123	Human novel human protein, NHP #23 - Homo sapiens, 576 aa. [WO200161016-A2, 23-AUG-2001]	143..674 31..574	224/564 (39%) 339/564 (59%)	e-109
AAU07119	Human novel human protein, NHP #19 - Homo sapiens, 560 aa. [WO200161016-A2, 23-AUG-2001]	143..654 31..554	213/544 (39%) 327/544 (59%)	e-102
AAU07115	Human novel human protein, NHP #15 - Homo sapiens, 520 aa. [WO200161016-A2, 23-AUG-2001]	143..606 31..495	196/481 (40%) 300/481 (61%)	5e-97

In a BLAST search of public sequence databases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9E.

Table 9E. Public BLASTP Results for NOV9a				
Protein Accession Number	Protein/Organism/Length	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9JLB2	PALS1 - Mus musculus (Mouse), 675 aa.	1..675 1..675	652/675 (96%) 665/675 (97%)	0.0

Q9H9Q0	CDNA FLJ12615 FIS, CLONE NT2RM4001629, WEAKLY SIMILAR TO MAGUK P55 SUBFAMILY MEMBER 3 - Homo sapiens (Human), 503 aa.	173..675 1..503	501/503 (99%) 501/503 (99%)	0.0
AAL40935	STARDUST PROTEIN MAGUK1 ISOFORM - Drosophila melanogaster (Fruit fly), 1289 aa.	252..674 829..1282	252/460 (54%) 327/460 (70%)	e-140
Q9W3H6	CG1617 PROTEIN - Drosophila melanogaster (Fruit fly), 794 aa.	252..674 294..787	252/500 (50%) 327/500 (65%)	e-132
Q9W7F1	P55-RELATED MAGUK PROTEIN DLG3 - Brachydanio rerio (Zebrafish) (Zebra danio), 576 aa.	142..673 30..573	209/556 (37%) 335/556 (59%)	e-105

PFam analysis predicts that the NOV9a protein contains the domains shown in the Table 9F.

Table 9F. Domain Analysis of NOV9a			
Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
L27: domain 1 of 1	186..238	19/56 (34%) 39/56 (70%)	0.049
PDZ: domain 1 of 1	256..335	21/83 (25%) 58/83 (70%)	9.7e-12
SH3: domain 1 of 1	348..415	19/68 (28%) 46/68 (68%)	0.026
Guanylate_kin: domain 1 of 1	515..624	54/113 (48%) 87/113 (77%)	6.2e-38
Peptidase_S15: domain 1 of 1	642..658	6/17 (35%) 13/17 (76%)	8.2
Caulimo_mov: domain 1 of 1	420..673	59/335 (18%) 156/335 (47%)	6.1

Example 10.

5   The NOV10 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 10A.

Table 10A. NOV10 Sequence Analysis		
	SEQ ID NO: 25	576 bp
NOV10a, CG58572-01 DNA Sequence	ACCTTACTAGAAAAATGAAACCTGATGAAACTCCTATGTTTGACCCAAGTCTACTCAA AGAAGTGGACTGGAGTCAGAATACAGCTACATTTTCTCCAGCCATTTCCCCAACACAT CCTGGAGAAGGCTTGGTTTTGAGGCCTCTTTGTACTGCTGACTTAAATAGAGGTTTTT TTAAGGTATTGGGTCAGCTAACAGAGACTGGAGTTGTCAGCCCTGAACAATTTATGGA ATCTTTTGAGCATATGAAGAAATCTGGGGATTATTATGTTACAGTTGTAGAAGATGTG ACTCTAGGACAGATTGTTGCTACGGCAACTCTGATTATAGAACATAAATTCATCCATT CCTGTGCTAAGAGAGGAAGAGTAGAAGATGTTGTTGTTAGTGATGAATGCAGAGGAAA GCAGCTTGGCAAATTGTTATTATCAACCCTTACTTTGCTAAGCAAGAAACTGAACTGT TACAAGATTACCCTTGAATGTCTACCACAAAATGTTGGTTTCTATAAAAGTTTGGAT ATACTGTATCTGAAGAAAACTACATGTGTCGGAGGTTTCTAAAGTAA<u>AAATCTT</u>	
	ORF Start: ATG at 15	ORF Stop: TAA at 567
	SEQ ID NO: 26	184 aa	MW at 20749.9kD
NOV10a, CG58572-01 Protein Sequence	MKPDETPMFDPSLLKEVDWSQNTATFSPAISPTHPGEGLVLRPLCTADLNRGFFKVLG QLTETGVVSPEQFMESFEHMKKSGDYYVTVVEDVTLGQIVATATLIIEHKFIHSCAKR GRVEDVVVSDECRGKQLGKLLLSTLTLLSKKLNCYKITLECLPQNVGFYKKFGYTVSE ENYMCRRFLK	
	SEQ ID NO: 27	560 bp
NOV10b, CG58572-02 DNA Sequence	ATGAAACCTGATGAAACTCCTATGTTTGACCCAAGTCTACTCAAAGAAGTGGACTGGA GTCAGAATACAGCTACATTTTCTCCAGCCATTTCCCCAACACATCCTGGAGAAGGCTT GGTTTTGGGGCCTCTTTGTACTGCTGACTTAAATAGAGGTTTTTTTAAGGTATTGGGT CAGCTAACAGAGACTGGAGTTGTCAGCCCTGAACAATTTATGAAATCTTTTGAGCATA TGAAGAAATCTGGGGATTATTATGTTACAGTTGTAGAAGATGTGACTCTAGGACAGAT TGTTGCTACGGCAACTCTGATTATAGAACATAAATTCATCCATTCCTGTGCTAAGAGA GGAAGAGTAGAAGATGTTGTTGTTAGTGATGAATGCAGAGGAAAGCAGCTTGGCAAAT TGTTATTATCAACCCTTACTTTGCTAAGCAAGAAACTGAACTGTTACAAGATTACCCT TGAATGTCTACCACAAAATGTTGGTTTCTATAAAAGTTTGGATATACTGTATCTGAA GAAAACTACATGTGTCGGAGGTTTCTAAAGTAA<u>AAATC</u>	
	ORF Start: ATG at 1	ORF Stop: TAA at 553
	SEQ ID NO: 28	184 aa	MW at 20649.8kD
NOV10b, CG58572-02 Protein Sequence	MKPDETPMFDPSLLKEVDWSQNTATFSPAISPTHPGEGLVLGPLCTADLNRGFFKVLG QLTETGVVSPEQFMKSFEHMKKSGDYYVTVVEDVTLGQIVATATLIIEHKFIHSCAKR GRVEDVVVSDECRGKQLGKLLLSTLTLLSKKLNCYKITLECLPQNVGFYKKFGYTVSE ENYMCRRFLK	

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 10B.

Table 10B. Comparison of NOV10a against NOV10b.		
Protein Sequence	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV10b	1..184 1..184	163/184 (88%) 164/184 (88%)

Further analysis of the NOV10a protein yielded the following properties shown in Table 10C.

Table 10C. Protein Sequence Properties NOV10a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.1206 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)

SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV10a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 10D.

Table 10D. Geneseq Results for NOV10a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG67123	Amino acid sequence of human 50287 transferase - Homo sapiens, 184 aa. [WO200164904-A2, 07-SEP-2001]	1..184 1..184	183/184 (99%) 184/184 (99%)	e-105
AAB73505	Human transferase HTFS-12, SEQ ID NO:12 - Homo sapiens, 184 aa. [WO200132888-A2, 10-MAY-2001]	1..184 1..184	183/184 (99%) 184/184 (99%)	e-105
AAB63700	Human gastric cancer associated antigen protein sequence SEQ ID NO:1062 - Homo sapiens, 200 aa. [WO200073801-A2, 07-DEC-2000]	1..184 17..200	183/184 (99%) 184/184 (99%)	e-105
AAU07779	Human novel transferase protein, NHP #22 - Homo sapiens, 184 aa. [WO200164903-A2, 07-SEP-2001]	1..184 1..184	182/184 (98%) 183/184 (98%)	e-104
AAM79992	Human protein SEQ ID NO 3638 - Homo sapiens, 206 aa. [WO200157190-A2, 09-AUG-2001]	1..184 23..206	181/184 (98%) 183/184 (99%)	e-104

5   In a BLAST search of public sequence databases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10E.

Table 10E. Public BLASTP Results for NOV10a				
Protein Accession Number	Protein/Organism/Length	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96EK6	SIMILAR TO GLUCOSAMINE-PHOSPHATE N-ACETYLTRANSFERASE - Homo sapiens (Human), 184 aa.	1..184 1..184	183/184 (99%) 184/184 (99%)	e-104
Q9JK38	EMEG32 PROTEIN (GLUCOSAMINE-PHOSPHATE N-ACETYLTRANSFERASE) - Mus musculus (Mouse), 184 aa.	1..184 1..184	180/184 (97%) 182/184 (98%)	e-102
Q9VAI0	Probable glucosamine-phosphate N-acetyltransferase (EC 2.3.1.4) (Phosphoglucosamine transacetylase) (Phosphoglucosamine acetylase) - Drosophila melanogaster (Fruit fly), 219 aa.	4..176 6..179	84/174 (48%) 123/174 (70%)	2e-43
Q17427	Probable glucosamine-phosphate N-acetyltransferase (EC 2.3.1.4) (Phosphoglucosamine transacetylase) (Phosphoglucosamine acetylase) - Caenorhabditis elegans, 165 aa.	32..182 15..165	65/152 (42%) 98/152 (63%)	1e-28
O45811	T23G11.2 PROTEIN - Caenorhabditis elegans, 347 aa.	42..184 201..340	63/143 (44%) 88/143 (61%)	3e-26

PFam analysis predicts that the NOV10a protein contains the domains shown in the Table 10F.

Table 10F. Domain Analysis of NOV10a			
Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Acetyltransf: domain 1 of 1	89..171	22/87 (25%) 62/87 (71%)	6.5e-13

Example 11.

5   The NOV11 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 11A.

126

Table 11A. NOV11 Sequence Analysis	
	SEQ ID NO: 29 | 709 bp
NOV11a, CG58564-01 DNA Sequence	CCCGCGGGCCAGCACCATGGAGGACGTGAAGCTGGAGTTCCCTTCCCTTCCACAGTGC AAGGAAGACGCCGAGGAGTGGACCTACCCTATGAGACGAGAGATGCAGGAAATTTTAC CTGGATTGTTCTTAGGCCCCATATTCATCTGCTATGAAAAGCAAGCTACCTGTACTACA GAAACATGGAATAACCCATATAATATGCATACGACAAAATATTGAAGCAAACTTTATT AAACCAAACTTTCAGCAGTTATTTAGGTATTTAGTCCTGGATATTGCAGATAATCCAG TTGAAAATATAATACGTTTTTTCCCTATGACTAAGGAATTTATTGATGGGAGCTTACA AATGGGAGGTAAAGTTCTTGTGCATGGAAATGCAGGGATCTCCAGAAGTGCAGCCTTT GTTATTGCATACATTATGGAAACATTTGGAATGAAGTACAGGGATGCTTTTGCTTATG TTCAAGAAGAAGATTTTGTATTAATCCTAATGCTGGATTTGTCCATCAACTTCAGGA ATATGAAGCCATCTACCTAGCAAAATTAACAATACAGATGATGTCACCACTCCAGATA GAAAGGTCATTATCTGTTCATTCTGGTACCACAGGTAGTTTGAAGAGAACACATGAAG AAGAGGATGATTTTGGAACCATGCAAGTGGCGACTGCACAGAATGGCTGACTTGAAGA GCAACATCATAGA
	ORF Start: ATG at 17 | ORF Stop: TGA at 686
	SEQ ID NO: 30 | 223 aa | MW at 25492.2kD
NOV11a, CG58564-01 Protein Sequence	MEDVKLEFPSLPQCKEDAEEWTYPMRREMQEILPGLFLGPYSSAMKSKLPVLQKHGIT HIICIRQNIEANFIKPNFQQLFRYLVLDIADNPVENIIRFFPMTKEFIDGSLQMGGKV LVHGNAGISRSAAFVIAYIMETFGMKYRDAFAYVQERRFCINPNAGFVHQLQEYEAIY LAKLTIQMMSPLQIERSLSVHSGTTGSLKRTHEEEDDFGTMQVATAQNG
	SEQ ID NO: 31 | 724 bp
NOV11b, CG58564-02 DNA Sequence	ACTCTCCCACCCCACCCACCAGAATGGCGGGCCAGCACCATGGAGGACGTGAAGCTGG AGTTCCCTTCCCTTCCACAGTGCAAGGAAGACGCCGAGGAGTGGACCTACCCTATGAG ACGAGAGATGCAGGAAATTTTATCTGGATTGTTCTTAGGCCCCATATTCATCTGCTATG AAAAGCAAGCTACCTGTACTACAGAAACATGGAATAACCCATATAATATGCATACGAC AAAATATTGAAGCAAACTTTATTAAACCAAACTTTCAGCAGTTATTTAGATATTTAGT CCTGGATATTGCAGATAATCCAGTTGAAAATATAATACGTTTTTTCCCTATGACTAAG GAATTTATTGATGGGAGCTTACAAATGGGAGGAAAAGTTCTTGTGCATGGAAATGCAG GGATCTCCAGAAGTGCAGCCTTTGTTATTGCATACATTATGGAAACATTTGGAATGAA GTACAGAGATGCTTTTGCTTATGTTCAAGAAGAAGATTTTGTATTAATCCTAATGCT GGATTTGTCCATCAACTTCAGGAATATGAAGCCATCTACCTAGCAAAATTAACAATAC AGATGATGTCACCACTCCAGATAGAAAGGTCATTATCTGTTCATTCTGGTACCACAGG CAGTTTGAAGAGAACACATGAAGAAGAGGATGATTTTGGAACCATGCAAGTGGCGACT GCACAGAATGGCTGACTTGAAGAGCAAC
	ORF Start: ATG at 40 | ORF Stop: TGA at 709
	SEQ ID NO: 32 | 223 aa | MW at 25482.1kD
NOV11b, CG58564-02 Protein Sequence	MEDVKLEFPSLPQCKEDAEEWTYPMRREMQEILSGLFLGPYSSAMKSKLPVLQKHGIT HIICIRQNIEANFIKPNFQQLFRYLVLDIADNPVENIIRFFPMTKEFIDGSLQMGGKV LVHGNAGISRSAAFVIAYIMETFGMKYRDAFAYVQERRFCINPNAGFVHQLQEYEAIY LAKLTIQMMSPLQIERSLSVHSGTTGSLKRTHEEEDDFGTMQVATAQNG
	SEQ ID NO: 33 | 545 bp
NOV11c, CG58564-03 DNA Sequence	ACTCTCCCACCCCACCCACCAGCCCGCGGGCCAGCACCATGGAGGACGTGAAGCTGGA GTTCCCTTCCCTTCCACAGTGCAAGGAAGACGCCGAGGAGTGGACCTACCCTATGAGA CGAGAGATGCAGGAAATTTTACCTGGATTGTTCTTAGGCCCCATATTCATCTGCTATGA AAAGCAAGCTACCTGTACTACAGAAACATTTGGAATGAAGTACAGAGATGCTTTTGCT TATGTTCAAGAAGAAGATTTTGTATTAATCCTAATGCTGGATTTGTCCATCAACTTC AGGAATATGAAGCCATCTACCTAGCAAAATTAACAATACAGATGATGTCACCACTCCA GATAGAAAGGTCATTATCTGTTCATTCTGGTACCACAGGCAGTTTGAAGAGAACACAT GAGGAAGAGGATGATTTTGGAACCATGCAAGTGGCGACTGCACAGAATGGCTGACTTG AAGAGCAACATCATAGAGTGTGAATTTCTATTTGGGAAGGAGAAAATACAAGAGAAAA TTATAATGTAAAATGGTAAAAAA
	ORF Start: ATG at 39 | ORF Stop: TGA at 210
	SEQ ID NO: 34 | 57 aa | MW at 6695.7kD
NOV11c, CG58564-03 Protein Sequence	MEDVKLEFPSLPQCKEDAEEWTYPMRREMQEILPGLFLGPYSSAMKSKLPVLQKHLE
	SEQ ID NO: 35 | 663 bp
NOV11d, CG58564-04 DNA Sequence	ACTCTCCCACCCCACCCACCAGCCCGCGGGCCAGCACCATGGAGGACGTGAAGCTGGA GTTCCCTTCCCTTCCACAGTGCAAGGAAGACGCCGAGGAGTGGACCTACCCTATGAGA CGAGAGATGCAGGAAATTTTACCTGGATTGTTCTTAGGCCCCATATTCATCTGCTATGA AAAGCAAGCTACCTGTACTACAGAAACATGGAATAACCCATATAATATGCATACGACA AAATATTGAAGCAAACTTTATTAAACCAAACTTTCAGCAGTTATTTAGACTAAGGAAT

127

	TTATTGATGGGAGCTTACAAATGGGAGGAAAAGTTCTTGTGCATGGAAATGCAGGGAT CTCCAGAAGTGCAGCCTTTGTTATTGCATACATTATGGAAACATTTGGAATGAAGTAC AGAGATGCTTTTGCTTATGTTCAAGAAAGAAGATTTTGTATTAATCCTAATGCTGGAT TTGTCCATCAACTTCAGGAATATGAAGCCATCTACCTAGCAAAATTAACAATACAGAT GATGTCACCACTCCAGATAGAAAGGTCATTATCTGTTCATTCTGGTACCACAGGCAGT TTGAAGAGAACACATGAAGAAGAGGATGATTTTGGAACCATGCAAGTGGCGACTGCAC AGAATGGCTGACTTGAAGAGCAACT
	ORF Start: ATG at 39 ORF Stop: TGA at 399
	SEQ ID NO: 36 120 aa MW at 14245.6kD
NOV11d, CG58564-04 Protein Sequence	MEDVKLEFPSLPQCKEDAEEWTYPMRREMQEILPGLFLGPYSSAMKSKLPVLQKHGIT HIICIRQNIEANFIKPNFQQLFRLRNLLMGAYKWEEKFLCMEMQGSPEVQPLLLHTLW KHLE

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 11B.

Table 11B. Comparison of NOV11a against NOV11b through NOV11d.		
Protein Sequence	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV11b	1..223 1..223	222/223 (99%) 222/223 (99%)
NOV11c	1..55 1..55	55/55 (100%) 55/55 (100%)
NOV11d	1..81 1..81	81/81 (100%) 81/81 (100%)

Further analysis of the NOV11a protein yielded the following properties shown in Table 11C.

5

Table 11C. Protein Sequence Properties NOV11a	
PSort analysis:	0.4698 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.1958 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11D.

Table 11D. Geneseq Results for NOV11a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU09017	Human dual specificity phosphatase 38692 - Homo sapiens, 223 aa. [WO200173059-A2, 04-OCT-2001]	1..223 1..223	223/223 (100%) 223/223 (100%)	e-128
AAE08552	Human phosphatase protein - Homo sapiens, 223 aa. [WO200160992-A2, 23-AUG-2001]	1..223 1..223	223/223 (100%) 223/223 (100%)	e-128
AAM41520	Human polypeptide SEQ ID NO 6451 - Homo sapiens, 236 aa. [WO200153312-A1, 26-JUL-2001]	1..223 14..236	223/223 (100%) 223/223 (100%)	e-128
AAM39734	Human polypeptide SEQ ID NO 2879 - Homo sapiens, 223 aa. [WO200153312-A1, 26-JUL-2001]	1..223 1..223	223/223 (100%) 223/223 (100%)	e-128
AAU23521	Novel human enzyme polypeptide #607 - Homo sapiens, 190 aa. [WO200155301-A2, 02-AUG-2001]	25..171 7..145	55/147 (37%) 80/147 (54%)	1e-18

In a BLAST search of public sequence databases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11E.

Table 11E. Public BLASTP Results for NOV11a				
Protein Accession Number	Protein/Organism/Length	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD10219	SEQUENCE 4 FROM PATENT WO0173059 - Homo sapiens (Human), 223 aa.	1..223 1..223	223/223 (100%) 223/223 (100%)	e-127
Q9DCF8	0610039A20RIK PROTEIN - Mus musculus (Mouse), 223 aa.	1..223 1..223	215/223 (96%) 221/223 (98%)	e-124
Q60970	PROTEIN TYROSINE PHOSPHATASE-LIKE - Mus musculus (Mouse), 223 aa.	1..223 1..223	214/223 (95%) 221/223 (98%)	e-124

Q60969	PROTEIN TYROSINE PHOSPHATASE-LIKE - Mus musculus (Mouse), 205 aa.	1..168 1..168	163/168 (97%) 167/168 (99%)	2e-93
Q99850	TYROSINE PHOSPHATASE-LIKE PROTEIN HOMOLOG HSTYXB - Homo sapiens (Human), 66 aa (fragment).	116..181 1..66	66/66 (100%) 66/66 (100%)	3e-31

PFam analysis predicts that the NOV11a protein contains the domains shown in the Table 11F.

Table 11F. Domain Analysis of NOV11a			
Pfam Domain	NOV11a Match Region	Identities/ Similarities for the Matched Region	Expect Value
DSPc: domain 1 of 1	28..173	64/172 (37%) 127/172 (74%)	2.2e-63
Y_phosphatase: domain 1 of 1	35..179	35/279 (13%) 93/279 (33%)	1.7

Example 12.

The NOV12 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis		
	SEQ ID NO: 37	3696 bp
NOV12a, CG57819-01 DNA Sequence	GTGTAAAAATACTGTCCATTTAATGTTTTCTGGGACTTTAGGTAAGAATATGAAAACT CAACCACCCTTGAGCAGGATGAACCGGGAGGAATTGGAGGACAGTTTCTTTCGACTTC GCGAAGATCACATGTTGGTGAAGGAGCTTTCTTGGAAGCAACAGGATGAGATCAAAAG GCTGAGGACCACCTTGCTGCGGTTGACCGCTGCTGGCCGGGACCTGCGGGTCGCGGAG GAGGCGGCGCCGCTCTCGGAGACCGCAAGGCGCGGGCAGAAGGCGGGATGGCGGCAGC GCCTCTCCATGCACCAGCGCCCCCAGATGCACCGACTGCAAGGGCATTTCCACTGCGT CGGCCCTGCCAGCCCCCGCCGCGCCCAGCCTCGCGTCCAAGTGGGACACAGACAGCTC CACACAGCCGGTGCACCGGTGCCGGAGAAACCCAAGAGGGGTAGGGACAGGCTGAGCT ACACAGCCCCTCCATCGTTTAAGGAGCATGCGACAAATGAAAACAGAGGTGAAGTAGC CAGTAAACCCAGTGAACTGGCCCACATCATGGCCAGCAATACCATGCAAGTGGAAGAG CCACCCAAGTCTCCTGAGAAAATGTGGCCTAAAGATGAAAATTTTGAACAGAGAAGCT CATTGGAGTGTGCTCAGAAGGCTGCAGAGCTTCGGGCTTCCATTAAAGAGAAGGTAGA GCTGATTCGACTTAAGAAGCTCTTACATGAAAGAAATGCTTCATTGGTTATGACAAAA GCACAATTAACAGAAGTTCAAGAGGTGAGTTGCCATCTTTTGACCCAGAATCAGGGAA TCCTGAGTGCAGCCCATGAGGCCCTCCTCAAGCAAGTGAATGAGCTCAGGGCAGAGCT GAAGGAAGAAAGCAAGAAGGCTGTGAGCTTGAAGAGCCAACTGGAAGATGTGTCTATC TTGCAGATGACTCTGAAGGAGTTTCAGGAGAGAGTTGAAGATTTGGAAAAAGAACGAA AATTGCTGAATGACAATTATGACAAACTCTTAGAAAGCAGTGACAGCTCCAGTCAGCC CCACTGGAGCAACGAGCTCATAGCGGAACAGCTACAGCAGCAAGTCTCTCAGCTGCAG GATCAGCTGGATGCTGAGCTGGAGGACAAGAGAAAAGTTTTACTTGAGCTGTCCAGGG AGAAAGCCCAAAATGAGGATCTGAAGCTTGAAGTCACCAACATACTTCAGAAGCATAA ACAGGAAGTAGAGCTCCTCCAAAATGCAGCCACAATTTCCCAACCTCCTGACAGGCAA TCTGAACCAGCCACTCACCCAGCTGTATTGCAAGAGAACACTCAGATCCAGCCAAGTG AACCCAAAAACCAAGAAGAAAGAAACTGTCCCAGGTGCTAAATGAGTTGCAAGTATC ACACGCAGAGACCACATTGGAACTAGAAAAGACCAGGGACATGCTTATTCTGCAGCGC AAAATCAACGTGTGTTATCAGGAGGAACTGGAGGCAATGATGACAAAAGCTGACAATG

130

	ATAATAGAGATCACAAAGAAAAGCTGGAGAGGTTGACTCGACTACTAGACCTCAAGAA TAACCGTATCAAGCAGCTGGAAGAACAGCTCAAAGATGTTGCTTATGGCACCCGACCG TTGTCGTTATGTTTGGAAACACTGCCAGCCCATGGAGATGAGGATAAAGTGGATATTT CTCTGCTGCATCAGGGTGAGAATCTTTTTGAACTGCACATCCACCAGGCCTTCCTGAC ATCTGCCGCCCTAGCTCAGGCTGGAGATACCCAACCTACCACTTTCTGCACCTATTCC TTCTATGACTTTGAAACCCACTGTACCCCATTATCTGTGGGGCCACAGCCCCTCTATG ACTTCACCTCCCAGTATGTGATGGAGACAGATTCGCTTTTCTTACACTACCTTCAAGA GGCTTCAGCCCGGCTTGACATACACCAGGCCATGGCCAGTGAACACAGCACTCTTGCT GCAGGATGGATTTGCTTTGACAGGGTGCTAGAGACTGTGGAGAAAGTCCATGGCTTGG CCACACTGATTGGTGCTGGTGGAGAAGAGTTCGGGGTTCTAGAGTACTGGATGAGGCT GCGTTTCCCCATAAAACCCAGCCTACAGGCGTGCAATAAACGAAAGAAAGCCCAGGTC TACCTGTCAACCGATGTGCTTGGAGGCCGGAAGGCCCAGGAAGAGGAGGTGAGATCGG AGTCTTGGGAACCTCAGAACGAGCTGTGGATTGAAATCACCAAGTGCTGTGGCCTCCG GAGTCGATGGCTGGGAACTCAACCCAGTCCATATGCTGTGTACCGCTTCTTCACCTTT TCTGACCATGACACTGCCATCATTCCAGCCAGTAACAACCCCTACTTTAGAGACCAGG CTCGATTCCCAGTGCTTGTGACCTCTGACCTGGACCATTATCTGAGACGGGAGGCCTT GTCTATACATGTTTTTGATGATGAAGACTTAGAGCCTGGCTCGTATCTTGGCCGAGCC CGAGTGCCTTTACTGCCTCTTGCAAAAAATGAATCTATCAAAGGTGATTTTAACCTCA CTGACCCTGCAGAGAAACCCAACGGATCTATTCAAGTGCAACTGGATTGGAAGTTTCC CTACATACCCCCTGAGAGCTTCCTGAAACCAGAAGCTCAGACTAAGGGGAAGGATACC AAGGACAGTTCAAAGATCTCATCTGAAGAGGAAAAGGCTTCATTTCCTTCCCAGGATC AGATGGCATCTCCTGAGGTTCCCATTGAAGCTGGCCAGTATCGATCTAAGAGAAAACC TCCTCATGGGGGAGAAAGAAAGGAGAAGGAGCACCAGGTTGTGAGCTACTCAAGAAGA AAACATGGCAAAAGAATAGGTGTTCAAGGAAAGAATAGAATGGAGTATCTTAGCCTTA ACATCTTAAATGGAAATACACTGAAGCAGGTGAATTACACTGAGTGGAAGTTCTCAGA GACTAACAGCTTCATAGGTGATGGCTTTAAAAATCAGCACGAGGAAGAGGAAATGACA TTATCCCATTCAGCACTGAAACAGAAGGAACCTCTACATCCTGTAAATGACAAAGAAT CCTCTGAACAAGGTTCTGAAGTCAGTGAAGCACAAACTACCGACAGTGATGATGTCAT AGTGCCACCCATGTCTCAGAAATATCCTAAGGCAGATTCAGAGAAGATGTGCATTGAA ATTGTCTCCCTGGCCTTCTACCCAGAGGCAGAAGTGATGTCTGATGAGAACATAAAC AGGTGTATGTGGAGTACAAATTCTACGACCTACCCTTGTCGGAGACAGAGACTCCAGT GTCCCTAAGGAAGCCTAGGGCAGGAGAAGAAATCCACTTTCACTTTAGCAAGGTAATA GACCTGGACCCACAGGAGCAGCAAGGCCGAAGGCGGTTTCTGTTCGACATGCTGAATG GACAAGATCCTGATCAAGGACAGTTAAAGTTTACAGTGGTAAGTGATCCTCTGGATGA AGAAAAGAAAGAATGTGAAGAAGTGGGATATGCATATCTTCAACTGTGGCAGATCCTG GAGTCAGGAAGAGATATTCTAGAGCAAGAGCTAGACGTTGTTAGCCCTGAAGATCTGG CTACCCCAATAGGAAGGCTGAAGGTTTCCCTTCAAGCAGCTGCTGTCCTCCATGCTAT TTACAAGGAGATGACTGAAGATTTGTTTTCATGAAGGAACAA
	ORF Start: ATG at 23	**ORF Stop: TGA at 3686**
	SEQ ID NO: 38	**1221 aa**	**MW at 139825.2kD**
NOV12a, CG57819-01 Protein Sequence	MFSGTLGKNMKTQPPLSRMNREELEDSFFRLREDHMLVKELSWKQQDEIKRLRTTLLR LTAAGRDLRVAEEAAPLSETARRGQKAGWRQRLSMHQRPQMHRLQGHFHCVGPASPRR AQPRVQVGHRQLHTAGAPVPEKPKRGRDRLSYTAPPSFKEHATNENRGEVASKPSELA HIMASNTMQVEEPPKSPEKMWPKDENFEQRSSLECAQKAAELRASIKEKVELIRLKKL LHERNASLVMTKAQLTEVQEVSCHLLTQNQGILSAAHEALLKQVNELRAELKEESKKA VSLKSQLEDVSILQMTLKEFQERVEDLEKERKLLNDNYDKLLESSDSSSQPHWSNELI AEQLQQQVSQLQDQLDAELEDKRKVLLELSREKAQNEDLKLEVTNILQKHKQEVELLQ NAATISQPPDRQSEPATHPAVLQENTQIQPSEPKNQEEKKLSQVLNELQVSHAETTLE LEKTRDMLILQRKINVCYQEELEAMMTKADNDNRDHKEKLERLTRLLDLKNNRIKQLE EQLKDVAYGTRPLSLCLETLPAHGDEDKVDISLLHQGENLFELHIHQAFLTSAALAQA GDTQPTTFCTYSFYDFETHCTPLSVGPQPLYDFTSQYVMETDSLFLHYLQEASARLDI HQAMASEHSTLAAGWICFDRVLETVEKVHGLATLIGAGGEEFGVLEYWMRLRFPIKPS LQACNKRKKAQVYLSTDVLGGRKAQEEEVRSESWEPQNELWIEITKCCGLRSRWLGTQ PSPYAVYRFFTFSDHDTAIIPASNNPYFRDQARFPVLVTSDLDHYLRREALSIHVFDD EDLEPGSYLGRARVPLLPLAKNESIKGDFNLTDPAEKPNGSIQVQLDWKFPYIPPESF LKPEAQTKGKDTKDSSKISSEEEKASFPSQDQMASPEVPIEAGQYRSKRKPPHGGERK EKEHQVVSYSRRKHGKRIGVQGKNRMEYLSLNILNGNTLKQVNYTEWKFSETNSFIGD GFKNQHEEEEMTLSHSALKQKEPLHPVNDKESSEQGSEVSEAQTTDSDDVIVPPMSQK YPKADSEKMCIEIVSLAFYPEAEVMSDENIKQVYVEYKFYDLPLSETETPVSLRKPRA GEEIHFHFSKVIDLDPQEQQGRRRFLFDMLNGQDPDQGQLKFTVVSDPLDEEKKECEE VGYAYLQLWQILESGRDILEQELDVVSPEDLATPIGRLKVSLQAAAVLHAIYKEMTED LFS

Further analysis of the NOV12a protein yielded the following properties shown in Table 12B.

Table 12B. Protein Sequence Properties NOV12a	
PSort analysis:	0.9600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 12C.

Table 12C. Geneseq Results for NOV12a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM78558	Human protein SEQ ID NO 1220 - Homo sapiens, 1179 aa. [WO200157190-A2, 09-AUG-2001]	63..1219 47..1177	400/1193 (33%) 640/1193 (53%)	e-172
AAM79542	Human protein SEQ ID NO 3188 - Homo sapiens, 1160 aa. [WO200157190-A2, 09-AUG-2001]	63..1219 28..1158	400/1193 (33%) 640/1193 (53%)	e-172
AAM41414	Human polypeptide SEQ ID NO 6345 - Homo sapiens, 1160 aa. [WO200153312-A1, 26-JUL-2001]	63..1219 28..1158	400/1193 (33%) 640/1193 (53%)	e-172
AAM39628	Human polypeptide SEQ ID NO 2773 - Homo sapiens, 1128 aa. [WO200153312-A1, 26-JUL-2001]	118..1219 47..1126	390/1138 (34%) 623/1138 (54%)	e-171
AAG75661	Human colon cancer antigen protein SEQ ID NO:6425 - Homo sapiens, 118 aa. [WO200122920-A2, 05-APR-2001]	445..523 33..111	40/79 (50%) 56/79 (70%)	1e-13

In a BLAST search of public sequence databases, the NOV12a protein was found to have homology to the proteins shown in the BLASTP data in Table 12D.

132

Table 12D. Public BLASTP Results for NOV12a				
Protein Accession Number	**Protein/Organism/Length**	**NOV12a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96KN7	RPGR-INTERACTING PROTEIN 1 - Homo sapiens (Human), 1286 aa.	7..1221 29..1286	1203/1258 (95%) 1207/1258 (95%)	0.0
Q96QA8	RPGR-INTERACTING PROTEIN 1 - Homo sapiens (Human), 1286 aa.	7..1221 29..1286	1203/1258 (95%) 1207/1258 (95%)	0.0
Q9GLM3	RPGR-INTERACTING PROTEIN-1 - Bos taurus (Bovine), 1221 aa.	1..1221 1..1221	922/1234 (74%) 1031/1234 (82%)	0.0
Q9NR40	RPGR-INTERACTING PROTEIN - Homo sapiens (Human), 902 aa.	331..1221 1..902	883/902 (97%) 888/902 (97%)	0.0
Q9HBK6	RPGR-INTERACTING PROTEIN-1 - Homo sapiens (Human), 762 aa.	471..1221 1..762	742/763 (97%) 746/763 (97%)	0.0

PFam analysis predicts that the NOV12a protein contains the domains shown in the Table 12E.

Table 12E. Domain Analysis of NOV12a			
Pfam Domain	**NOV12a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
PFEMP: domain 1 of 1	293..413	23/176 (13%) 82/176 (47%)	7.9
C2: domain 1 of 1	736..825	14/101 (14%) 54/101 (53%)	1.4

Example 13.

5   The NOV13 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 13A.

133

Table 13A. NOV13 Sequence Analysis		
	SEQ ID NO: 39	678 bp
NOV13a, CG57789-01 DNA Sequence	TGGGGCGGGAGGCATGGTCTCCACCTACCGGGTGGCCGTGCTGGGGGCGCGAGGTGTG GGCAAGAGTGCCATCGTGCGCCAGTTCTTGTACAACGAGTTCAGCGAGGTCTGCGTCC CCACCACCGCCCGCCGCCTTTACCTGCCTGCTGTCGTCATGAACGGCCACGTGCACGA CCTCCAGATCCTCGACTTTCCACCCATCAGCGCCTTCCCTGTCAATACGCTCCAGGAG TGGGCAGACACCTGCTGCAGGGGACTCCGGAGTGTCCACGCCTACATCCTGGTCTACG ACATCTGCTGCTTTGACAGCTTTGAGTACGTCAAGACCATCCGCCAGCAGATCCTGGA GACGAGGGTGATCGGAACCTCAGAGACGCCCATCATCATCGTGGGCAACAAGCGGGAC CTGCAGCGCGGACGCGTGATCCCGCGCTGGAACGTGTCGCACCTGGTACGCAAGACCT GGAAGTGCGGCTACGTGGAATGCTCGGCCAAGTACAACTGGCACATCCTGCTGCTCTT CAGCGAGCTGCTCAAGAGCGTCGGCTGCGCCCGTTGCAAGCACGTGCACGCTGCCCTG CGCTTCCAGGGCGCGCTGCGCCGCAACCGCTGCGCCATCATGTGACGCCTGCGCGCCC CTCGGGCTGCACCGGCACTGGCCGAGCGGAGGGCGGGGCC	
	ORF Start: ATG at 14	ORF Stop: TGA at 623
	SEQ ID NO: 40	203 aa	MW at 23229.0kD
NOV13a, CG57789-01 Protein Sequence	MVSTYRVAVLGARGVGKSAIVRQFLYNEFSEVCVPTTARRLYLPAVVMNGHVHDLQIL DFPPISAFPVNTLQEWADTCCRGLRSVHAYILVYDICCFDSFEYVKTIRQQILETRVI GTSETPIIIVGNKRDLQRGRVIPRWNVSHLVRKTWKCGYVECSAKYNWHILLLFSELL KSVGCARCKHVHAALRFQGALRRNRCAIM	
	SEQ ID NO: 41	682 bp
NOV13b, CG57789-02 DNA Sequence	TGGGAGGCATGGTCTCCACCTACCGGGTGGCCGTGCTGGGGGCGCGAGGTGTGGGCAA GAGTGCCATCGTGCGCCAGTTCTTGTACAACGAGTTCAGCGAGGTCTGCGTCCCCACC ACCGCCCGCCGCCTTTACCTGCCTGCTGTCGTCATGAACGGCCACGTGCACGACCTCC AGATCCTCGACTTTCCACCCATCAGCGCCTTCCCTGTCAATACGCTCCAGGAGTGGGC AGACACCTGCTGCAGGGGACTCCGGAGTGTCCACGCCTACATCCTGGTCTACGACATC TGCTGCTTTGACAGCTTTGAGTACGTCAAGACCATCCGCCAGCAGATCCTGGAGACGA GGGTGATCGGAACCTCAGAGACGCCCATCATCATCGTGGGCAACAAGCGGGACCTGCA GCGCGGACGCGTGATCCCGCGCTGGAACGTGTCGCACCTGGTACGCAAGACCTGGAAG TGCGGCTACGTGGAATGCTCGGCCAAGTACAACTGGCACATCCTGCTGCTCTTCAGCG AGCTGCTCAAGAGCGTCGGCTGCGCCCGTTGCAAGCACGTGCACGCTGCCCTGCGCTT CCAGGGCGCGCTGCGCCGCAACCGCTGCGCCATCATGTGACGCCTGCGCGCCCCTCGG GCTGCACCGGCACTGGCCGAGCGGAGGGCACTGGCCGAGCGGAG	
	ORF Start: ATG at 9	ORF Stop: TGA at 618
	SEQ ID NO: 42	203 aa	MW at 23229.0kD
NOV13b, CG57789-02 Protein Sequence	MVSTYRVAVLGARGVGKSAIVRQFLYNEFSEVCVPTTARRLYLPAVVMNGHVHDLQIL DFPPISAFPVNTLQEWADTCCRGLRSVHAYILVYDICCFDSFEYVKTIRQQILETRVI GTSETPIIIVGNKRDLQRGRVIPRWNVSHLVRKTWKCGYVECSAKYNWHILLLFSELL KSVGCARCKHVHAALRFQGALRRNRCAIM	

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 13B.

Table 13B. Comparison of NOV13a against NOV13b.		
Protein Sequence	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV13b	1..203 1..203	203/203 (100%) 203/203 (100%)

Further analysis of the NOV13a protein yielded the following properties shown in Table 13C.

5

Table 13C. Protein Sequence Properties NOV13a	
PSort analysis:	0.6500 probability located in plasma membrane; 0.5064 probability located in mitochondrial matrix space; 0.3844 probability located in microbody (peroxisome); 0.2556 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 13D.

Table 13D. Geneseq Results for NOV13a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB42840	Human ORFX ORF2604 polypeptide sequence SEQ ID NO:5208 - Homo sapiens, 136 aa. [WO200058473-A2, 05-OCT-2000]	1..136 1..136	136/136 (100%) 136/136 (100%)	2e-75
AAM41682	Human polypeptide SEQ ID NO 6613 - Homo sapiens, 206 aa. [WO200153312-A1, 26-JUL-2001]	5..174 15..173	66/171 (38%) 89/171 (51%)	4e-18
AAM39896	Human polypeptide SEQ ID NO 3041 - Homo sapiens, 199 aa. [WO200153312-A1, 26-JUL-2001]	5..174 8..166	66/171 (38%) 89/171 (51%)	4e-18
AAY99656	Human GTPase associated protein-7 - Homo sapiens, 281 aa. [WO200031263-A2, 02-JUN-2000]	5..173 25..191	59/179 (32%) 87/179 (47%)	3e-14
AAR05075	RAP1A Gene product incorporating at least one peptide associated with ras oncogene - Synthetic, 184 aa. [WO9000179-A, 11-JAN-1990]	5..177 4..165	57/175 (32%) 90/175 (50%)	5e-14

In a BLAST search of public sequence databases, the NOV13a protein was found to

5 have homology to the proteins shown in the BLASTP data in Table 13E.

Table 13E. Public BLASTP Results for NOV13a				
Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96S79	RAS-LIKE PROTEIN/VTS58635 - Homo sapiens (Human), 203 aa.	1..203 1..203	203/203 (100%) 203/203 (100%)	e-118
Q92737	Ras-like protein RRP22 (RAS-related protein on chromosome 22) - Homo sapiens (Human), 203 aa.	1..203 1..203	105/204 (51%) 134/204 (65%)	3e-50
Q95KD9	HYPOTHETICAL 22.5 KDA PROTEIN - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 199 aa.	5..174 8..166	66/171 (38%) 89/171 (51%)	1e-17
Q96HU8	SIMILAR TO CG8500 GENE PRODUCT - Homo sapiens (Human), 199 aa.	5..174 8..166	66/171 (38%) 89/171 (51%)	1e-17
Q9NF75	EG:BACR37P7.8 PROTEIN - Drosophila melanogaster (Fruit fly), 306 aa.	5..174 48..210	61/174 (35%) 88/174 (50%)	4e-16

PFam analysis predicts that the NOV13a protein contains the domains shown in the Table 13F.

Table 13F. Domain Analysis of NOV13a			
Pfam Domain	NOV13a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Semialdhyde_dh: domain 1 of 1	4..14	4/11 (36%) 11/11 (100%)	0.75
ras: domain 1 of 1	6..203	56/224 (25%) 125/224 (56%)	1.2e-12

Example 14.

The NOV14 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis		
	SEQ ID NO: 43	1790 bp
NOV14a, CG57758-01 DNA Sequence	TCTCCCTCCCGCGCGATGGCCTCGGCGCTGAGCTATGTCTCCAAGTTCAAGTCCTTCG TGATCTTGTTCGTCACCCCGCTCCTGCTGCTGCCACTCGTCATTCTGATGCCCGCCAA GGTCAGTTGTGCCTACGTCATCATCCTCATGGCCATTTACTGGTGCACAGAAGTCATC CCTCTGGCTGTCACCTCTCTCATGCCTGTCTTGCTTTTCCCACTCTTCCAGATTCTGG ACTCCAGGCAGGTGTGTGTCCAGTACATGAAGGACACCAACATGCTGTTCCTGGGCGG CCTCATCGTGGCCGTGGCTGTGGAGCGCTGGAACCTGCACAAGAGGATCGCCCTGCGC ACGCTCCTCTGGGTGGGGGCCAAGCCTGCACGGCTGATGCTGGGCTTCATGGGCGTCA CAGCCCTCCTGTCCATGTGGATCAGTAACACGGCAACCACGGCCATGATGGTGCCCAT CGTGGAGGCCATATTGCAGCAGATGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTG GAGCTGGTGGACAAGGGCAAGGCCAAGGAGCTGCCAGGGAGTCAAGTGATTTTTGAAG GCCCCACTCTGGGGCAGCAGGAAGACCAAGAGCGGAAGAGGTTGTGTAAGGCCATGAC CCTGTGCATCTGCTACGCGGCCAGCATCGGGGGCACCGCCACCCTGACCGGGACGGGA CCCAACGTGGTGCTCCTGGGCCAGATGAACGAGTTGTTTCCTGACAGCAAGGACCTCG TGAACTTTGCTTCCTGGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTT CGCCTGGCTGTGGCTCCAGTTTGTTTACATGTTCTCCAGTTTTAAAAAGTCCTGGGGC TGCGGGCTAGAGAGCAAGAAAAACGAGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGT ACCGGAAGCTGGGGCCCTTGTCCTTCGCGGAGATCAACGTGCTGATCTGCTTCTTCCT GCTGGTCATCCTGTGGTTCTCCCGAGACCCCGGCTTCATGCCCGGCTGGCTGACTGTT GCCTGGGTGGAGGGTGAGACAAAGTATGTCTCCGATGCCACTGTGGCCATCTTTGTGG CCACCCTGCTATTCATTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGCAGCCAGAC TGAGGAAGGTAAGTCTCCTGTTCTGATCGCCCCCCCTCCCCTGCTGGATTGGAAGGTA ACCCAGGAGAAAGTGCCCTGGGGCATCGTGCTGCTACTAGGGGGCGGATTTGCTCTGG CTAAAGGATCCGAGGCCTCGGGGCTGTCCGTGTGGATGGGGAAGCAGATGGAGCCCTT GCACGCAGTGCCCCCGGCAGCCATCACCTTGATCTTGTCCTTGCTCGTTGCCGTGTTC ACTGAGTGCACAAGCAACGTGGCCACCACCACCTTGTTCCTGCCCATCTTTGCCTCCA TGTCTCGCTCCATCGGCCTCAATCCGCTGTACATCATGCTGCCCTGTACCCTGAGTGC CTCCTTTGCCTTCATGTTGCCTGTGGCCACCCCTCCAAATGCCATCGTGTTCACCTAT GGGCACCTCAAGGTTGCTGACATGGTGAAAACAGGAGTCATAATGAACATAATTGGAG TCTTCTGTGTGTTTTTGGCTGTCAACACCTGGGGACGGGCCATATTTGACTTGGATCA TTTCCCTGACTGGGCTAATGTGACACATATTGAGACTTAGGAAGAGCCACAAGACCAC ACACACAGCCCTTACCCTCCTCAGGACTACCGAACCTTCTGGCACACCTT
	ORF Start: ATG at 16	ORF Stop: TAG at 1720
	SEQ ID NO: 44	568 aa	MW at 62592.9kD
NOV14a, CG57758-01 Protein Sequence	MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKVSCAYVIILMAIYWCTEVIPLAVT SLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLWV GAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDK GKAKELPGSQVIFEGPTLGQQEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVVL LGQMNELFPDSKDLVNFASWFAFAFPNMLVMLLFAWLWLQFVYMFSSFKKSWGCGLES KKNEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEG ETKYVSDATVAIFVATLLFIVPSQKPKFNFRSQTEEGKSPVLIAPPPLLDWKVTQEKV PWGIVLLLGGGFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTS NVATTTLFLPIFASMSRSIGLNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKV ADMVKTGVIMNIIGVFCVFLAVNTWGRAIFDLDHFPDWANVTHIET	
	SEQ ID NO: 45	1899 bp
NOV14b, CG57758-02 DNA Sequence	CGTCTCGCCCGCCAGTCTCCCTCCCGCGCGATGGCCTCGGCGCTGAGCTATGTCTCCA AGTTCAAGTCCTTCGTGATCTTGTTCGTCACCCCGCTCCTGCTGCTGCCACTCGTCAT TCTGATGCCCGCCAAGGTCAGTTGCTGTGCCTACGTCATCATCCTCATGGCCATTTAC TGGTGCACAGAAGTCATCCCTCTGGCTGTCACCTCTCTCATGCCTGTCTTGCTTTTCC CACTCTTCCAGATTCTGGACTCCAGGCAGGTGTGTGTCCAGTACATGAAGGACACCAA CATGCTGTTCCTGGGCGGCCTCATCGTGGCCGTGGCTGTGGAGCGCTGGAACCTGCAC AAGAGGATCGCCCTGCGCACGCTCCTCTGGGTGGGGGCCAAGCCTGCACGGCTGATGC TGGGCTTCATGGGCGTCACAGCCCTCCTGTCCATGTGGATCAGTAACACGGCAACCAC GGCCATGATGGTGCCCATCGTGGAGGCCATATTGCAGCAGATGGAAGCCACAAGCGCA GCCACCGAGGCCGGCCTGGAGGGACAAGGTACCACAATAAACAACCTGAATGCACTGG AGGATGATACAGTGAAAGCAGTACTAGGAGGAAAGTGTGTAGCTATAATAAGCACTTA CGTCAAAAAAGTAGAAAAACTTCAAATAAACAATCTAATGACACCTCTTAAAAAACTA GAAAAGCAAGAGCAACAGGACCTAGGGCCTGGCATCAGGCCTCAGGACTCTGCCCAGT GCCAGGAAGACCAAGAGCGGAAGAGGTTGTGTAAGGCCATGACCCTGTGCATCTGCTA CGCGGCCAGCATCGGGGGCACCGCCACCCTGACCGGGACGGGACCCAACGTGGTGCTC CTGGGCCAGATGAACGAGTTGTTTCCTGACAGCAAGGACCTCGTGAACTTTGCTTCCT GGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTTCGCCTGGCTGTGGCT CCAGTTTGTTTACATGTTCTCCAGTTTTAAAAAGTCCTGGGGCTGCGGGCTAGAGAGC AAGAAAAACGAGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTACCGGAAGCTGGGGC CCTTGTCCTTCGCGGAGATCAACGTGCTGATCTGCTTCTTCCTGCTGGTCATCCTGTG TGTTCTCCCGAGACCCCGGCTTCATGCCCGGCTGGCTGACTGTTGCCTGGGTGGAGGGT GAGACAAAGTCAGTCTCCGATGCCACTGTGGCCATCTTTGTGGCCACCCTGCTATTCA TTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGCAGCCAGACTGAGGAAGGTAAGTC TCCTGTTCTGATCGCCCCCCCTCCCCTGCTGGATTGGAAGGTAACCCAGGAGAAAGTG CCCTGGGGCATCGTGCTGCTACTAGGGGGCGGATTTGCTCTGGCTAAAGGATCCGAGG

	CCTCGGGGCTGTCCGTGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCCC GGCAGCCATCACCTTGATCTTGTCCTTGCTCGTTGCCGTGTTCACTGAGTGCACAAGC AACGTGGCCACCACCACCTTGTTCCTGCCCATCTTTGCCTCCATGTCTCGCTCCATCG GCCTCAATCCGCTGTACATCATGCTGCCCTGTACCCTGAGTGCCTCCTTTGCCTTCAT GTTGCCTGTGGCCACCCCTCCAAATGCCATCGTGTTCACCTATGGGCACCTCAAGGTT GCTGACATGGTAAAAACAGGAGTCATAATGAACATAATTGGAGTCTTCTGTGTGTTTT TGGCTGTCAACACCTGGGGACGGGCCATATTTGACTTGGATCATTTCCCTGACTGGGC TAATGTGACACATATTGAGACTTAGGAAGAGCCACAAGACCAC

	ORF Start: ATG at 31	ORF Stop: TAG at 1879	
	SEQ ID NO: 46	616 aa	MW at 67816.9kD

NOV14b, CG57758-02 Protein Sequence	MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKVSCCAYVIILMAIYWCTEVIPLAV TSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLW VGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLEGQG TTINNLNALEDDTVKAVLGGKCVAIISTYVKKVEKLQINNLMTPLKKLEKQEQQDLGP GIRPQDSAQCQEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPD SKDLVNFASWFAFAFPNMLVMLLFAWLWLQFVYMFSSFKKSWGCGLESKKNEKAALKV LQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKSVSDATV AIFVATLLFIVPSQKPKFNFRSQTEEGKSPVLIAPPPLLDWKVTQEKVPWGIVLLLGG GFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLP IFASMSRSIGLNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIM NIIGVFCVFLAVNTWGRAIFDLDHFPDWANVTHIET

	SEQ ID NO: 47	1899 bp

NOV14c, CG57758-03 DNA Sequence	CGTCTCGCCCGCCAGTCTCCCTCCCGCGCGATGGCCTCGGCGCTGAGCTATGTCTCCA AGTTCAAGTCCTTCGTGATCTTGTTCGTCACCCCGCTCCTGCTGCTGCCCACTCGTCAT TCTGATGCCCGCCAAGGTCAGTTGCTGTGCCTACGTCATCATCCTCATGGCCATTTAC TGGTGCACAGAAGTCATCCCTCTGGCTGTCACCTCTCTCATGCCTGTCTTGCTTTTCC CACTCTTCCAGATTCTGGACTCCAGGCAGGTGTGTGTCCAGTACATGAAGGACACCAA CATGCTGTTCCTGGGCGGCCTCATCGTGGCCGTGGCTGTGGAGCGCTGGAACCTGCAC AAGAGGATCGCCCTGCGCACGCTCCTCTGGGTGGGGGCCAAGCCTGCACGGCTGATGC TGGGCTTCATGGGCGTCACAGCCCTCCTGTCCATGTGGATCAGTAACACGGCAACCAC GGCCATGATGGTGCCCATCGTGGAGGCCATATTGCAGCAGATGGAAGCCACAAGCGCA GCCACCGAGGCCGGCCTGGAGGGACAAGGTACCACAATAAACAACCTGAATGCACTGG AGGATGATACAGTGAAAGCAGTACTAGGAGGAAAGTGTGTAGCTATAATAAGCACTTA CGTCAAAAAAGTAGAAAAACTTCAAATAAACAATCTAATGACACCTCTTAAAAAACTA GAAAAGCAAGAGCAACAGGACCTAGGGCCTGGCATCAGGCCTCAGGACTCTGCCCAGT GCCAGGAAGACCAAGAGCGGAAGAGGTTGTGTAAGGCCATGACCCTGTGCATCTGCTA CGCGGCCAGCATCGGGGGCACCGCCACCCTGACCGGGACGGGACCCAACGTGGTGCTC CTGGGCCAGATGAACGAGTTGTTTCCTGACAGCAAGGACCTCGTGAACTTTGCTTCCT GGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTTCGCCTGGCTGTGGCT CCAGTTTGTTTACATGTTCTCCAGTTTTAAAAAGTCCTGGGGCTGCGGGCTAGAGAGC AAGAAAAACGAGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTACCGGAAGCTGGGGC CCTTGTCCTTCGCGGAGATCAACGTGCTGATCTGCTTCTTCCTGCTGGTCATCCTGTG GTTCTCCCGAGACCCCGGCTTCATGCCCGGCTGGCTGACTGTTGCCTGGGTGGAGGGT GAGACAAAGTCAGTCTCCGATGCCACTGTGGCCATCTTTGTGGCCACCCTGCTATTCA TTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGCAGCCAGACTGAGGAAGGTAAGTC TCCTGTTCTGATCGCCCCCCCTCCCCTGCTGGATTGGAAGGTAACCCAGGAGAAAGTG CCCTGGGGCATCGTGCTGCTACTAGGGGGCGGATTGCTCTGGCTAAAGGATCCGAGG CCTCGGGGCTGTCCGTGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCCC GGCAGCCATCACCTTGATCTTGTCCTTGCTCGTTGCCGTGTTCACTGAGTGCACAAGC AACGTGGCCACCACCACCTTGTTCCTGCCCATCTTTGCCTCCATGTCTCGCTCCATCG GCCTCAATCCGCTGTACATCATGCTGCCCTGTACCCTGAGTGCCTCCTTTGCCTTCAT GTTGCCTGTGGCCACCCCTCCAAATGCCATCGTGTTCACCTATGGGCACCTCAAGGTT GCTGACATGGTAAAAACAGGAGTCATAATGAACATAATTGGAGTCTTCTGTGTGTTTT TGGCTGTCAACACCTGGGGACGGGCCATATTTGACTTGGATCATTTCCCTGACTGGGC TAATGTGACACATATTGAGACTTAGGAAGAGCCACAAGACCAC

	ORF Start: ATG at 31	ORF Stop: TAG at 1879	
	SEQ ID NO: 48	616 aa	MW at 67816.9kD

NOV14c, CG57758-03 Protein Sequence	MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKVSCCAYVIILMAIYWCTEVIPLAV TSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLW VGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLEGQG TTINNLNALEDDTVKAVLGGKCVAIISTYVKKVEKLQINNLMTPLKKLEKQEQQDLGP GIRPQDSAQCQEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPD SKDLVNFASWFAFAFPNMLVMLLFAWLWLQFVYMFSSFKKSWGCGLESKKNEKAALKV LQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKSVSDATV AIFVATLLFIVPSQKPKFNFRSQTEEGKSPVLIAPPPLLDWKVTQEKVPWGIVLLLGG GFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLP IFASMSRSIGLNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIM NIIGVFCVFLAVNTWGRAIFDLDHFPDWANVTHIET

	SEQ ID NO: 49	1606 bp

138

| NOV14d, CG57758-04 DNA Sequence | GATGGCCTCGGCGCTGAGCTATGTCTCCAAGTTCAAGTCCTTCGTGATCTTGTTCGTC ACCCCGCTCCTGCTGCTGCCACTCGTCATTCTGATGCCCGCCAAGTTTGTCAGGTGTG CCTACGTCATCATCCTCATGGCCATTTACTGGTGCACAGAAGTCATCCCTCTGGCTGT CACCTCTCTCATGCCTGTCTTGCTTTTCCCACTCTTCCAGATTCTGGACTCCAGGCAG GTGTGTGTCCAGTACATGAAGGACACCAACATGCTGTTCCTGGGCGGCCTCATCGTGG CCGTGGCTGTGGAGCGCTGGAACCTGCACAAGAGGATCGCCCTGCGCACGCTCCTCTG GGTGGGGGCCAAGCCTGCACGGCTGATGCTGGGCTTCATGGGCGTCACAGCCCTCCTG TCCATGTGGATCAGTAACACGGCAACCACGGCCATGATGGTGCCCATCGTGGAGGCCA TATTGCAGCAGATGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGA CAAGGGCAAGGCCAAGGAGCTGCCAGGGAGTCAAGTGATTTTTGAAGGCCCCACTCTG GGGCAGCAGGAAGACCAAGAGCGGAAGAGGTTGTGTAAGGCCATGACCCTGTGCATCT GCTACGCGGCCAGCATCGGGGGCACCGCCACCCTGACCGGGACGGGACCCAACGTGGT GCTCCTGGGCCAGATGAACGAGTTGTTTCCTGACAGCAAGGACCTCGTGAACTTTGCT TCCTGGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTTCGCCTGGCTGT GGCTCCAGTTTGTTTACATGAGATTCAATTTTAAAAAGTCCTGGGGCTGCGGGCTAGA GAGCAAGAAAAACGAGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTACCGGAAGTTG GGGCCCTTGTCCTTCGCGGAGATCAACGTGCTGATCTGCTTCTTCCTGCTGGTCATCC TGTGGTTCTCCCGAGACCCCGGCTTCATGCCCGGCTGGCTGACTGTTGCCTGGGTGGA GGGTGAGACAAAGTATGTCTCCGATGCCACTGTGGCCATCTTTGTGGCCACCCTGCTA TTCATTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGCAGCCAGACTGAGGAAGAAA GGAAAACTCCATTTTATCCCCCTCCCCTGCTGGATTGGAAGGTAACCCAGGAGAAAGT GCCCTGGGGCATCGTGCTGCTACTAGGGGGCGGATTTGCTCTGGCTAAAGGATCCGAG GCCTCGGGGCTGTCCGTGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCC CGGCAGCCATCACCTTGATCTTGTCCTTGCTCGTTGCCGTGTTCACTGAGTGCACAAG CAACGTGGCCACCACCACCTTGTTCCTGCCCATCTTTGCCTCCATGGTGAAAACAGGA GTCATAATGAACATAATTGGAGTCTTCTGTGTGTTTTTGGCTGTCAACACCTGGGGAC GGGCCATATTTGACTTGGATCATTTCCCTGACTGGGCTAATGTGACACATATTGAGAC **TTAG**GAAGAGCCACAAGACCACACACATAGCCCTTACCCT |

	ORF Start: ATG at 2	ORF Stop: TAG at 1568	
	SEQ ID NO: 50	522 aa	MW at 58109.6kD

| NOV14d, CG57758-04 Protein Sequence | MASALSYVSKFKSFVILFVTPLLLLPLVIILMPAKFVRCAYVIILMAIYWCTEVIPLAV TSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLW VGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVD KGKAKELPGSQVIFEGPTLGQQEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVV LLGQMNELFPDSKDLVNFASWFAFAFPNMLVMLLFAWLWLQFVYMRFNFKKSWGCGLE SKKNEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVE GETKYVSDATVAIFVATLLFIVPSQKPKFNFRSQTEEERKTPFYPPPLLDWKVTQEKV PWGIVLLLGGGFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTS NVATTTLFLPIFASMVKTGVIMNIIGVFCVFLAVNTWGRAIFDLDHFPDWANVTHIET |

	SEQ ID NO: 51	1781 bp

| NOV14e, CG57758-05 DNA Sequence | GATGGCCTCGGCGCTGAGCTATGTCTCCAAGTTCAAGTCCTTCGTGATCTTGTTCGTC ACCCCGCTCCTGCTGCTGCCACTCGTCATTCTGATGCCCGCCAAGTTTGTCAGGTGTG CCTACGTCATCATCCTCATGGCCATTTACTGGTGCACAGAAGTCATCCCTCTGGCTGT CACCTCTCTCATGCCTGTCTTGCTTTTCCCACTCTTCCAGATTCTGGACTCCAGGCAG GTGTGTGTCCAGTACATGAAGGACACCAACATGCTGTTCCTGGGCGGCCTCATCGTGG CCGTGGCTGTGGAGCGCTGGAACCTGCACAAGAGGATCGCCCTGCGCACGCTCCTCTG GGTGGGGGCCAAGCCTGCACGGCTGATGCTGGGCTTCATGGGCGTCACAGCCCTCCTG TCCATGTGGATCAGTAACACGGCAACCACGGCCATGATGGTGCCCATCGTGGAGGCCA TATTGCAGCAGATGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGA CAAGGGCAAGGCCAAGGAGCTGCCAGGGAGTCAAGTGATTTTTGAAGGCCCCACTCTG GGGCAGCAGGAAGACCAAGAGCGGAAGAGGTTGTGTAAGGCCATGACCCTGTGCATCT GCTACGCGGCCAGCATCGGGGGCACCGCCACCCTGACCGGGACGGGACCCAACGTGGT GCTCCTGGGCCAGATGAACGAGTTGTTTCCTGACAGCAAGGACCTCGTGAACTTTGCT TCCTGGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTTCGCCTGGCTGT GGCTCCAGTTTGTTTACATGAGATTCAATTTTAAAAAGTCCTGGGGCTGCGGGCTAGA GAGCAAGAAAAACGAGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTACCGGAAGTTG GGGCCCTTGTCCTTCGCGGAGATCAACGTGCTGATCTGCTTCTTCCTGCTGGTCATCC TGTGGTTCTCCCGAGACCCCGGCTTCATGCCCGGCTGGCTGACTGTTGCCTGGGTGGA GGGTGAGACAAAGTATGTCTCCGATGCCACTGTGGCCATCTTTGTGGCCACCCTGCTA TTCATTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGCAGCCAGACTGAGGAAGAAA GGAAAACTCCATTTTATCCCCCTCCCCTGCTGGATTGGAAGGTAACCCAGGAGAAAGT GCCCTGGGGCATCGTGCTGCTACTAGGGGGCGGATTTGCTCTGGCTAAAGGATCCGAG GCCTCGGGGCTGTCCGTGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCC CGGCAGCCATCACCTTGATCTTGTCCTTGCTCGTTGCCGTGTTCACTGAGTGCACAAG CAACGTGGCCACCACCACCTTGTTCCTGCCCATCTTTGCCTCCATGAATCACGTCCCC AAGAGCTTCTGTGTTCTGTACGGTGATGTTGCAGTGCTGTCTTTCCGCAGTCTCGCTC CATCGGCCTCAATCCGCTGTACATCATGCTGCCCTGTACCC**TGA**GTGCCTCCTTTGCC TTCATGTTGCCTGTGGCCACCCCTCCAAATGCCATCGTGTTCACCTATGGGCACCTCA AGGTTGCTGACATGGTGAAAACAGGAGTCATAATGAACATAATTGGAGTCTTCTGTGT GTTTTTGGCTGTCAACACCTGGGGACGGGCCATATTTGACTTGGATCATTTCCCTGAC TGGGCTAATGTGACACATATTGAGACTTAGGAAGAGCCACA |

	ORF Start: ATG at 2	ORF Stop: TGA at 1550	

	SEQ ID NO: 52	516 aa	MW at 57173.5kD
NOV14e, CG57758-05 Protein Sequence	MASALSYVSKFKSFVILFVTPLLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIPLAV TSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLW VGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVD KGKAKELPGSQVIFEGPTLGQQEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVV LLGQMNELFPDSKDLVNFASWFAFAFPNMLVMLLFAWLWLQFVYMRFNFKKSWGCGLE SKKNEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVE GETKYVSDATVAIFVATLLFIVPSQKPKFNFRSQTEEERKTPFYPPPLLDWKVTQEKV PWGIVLLLGGGFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTS NVATTTLFLPIFASMNHVPKSFCVLYGDVAVLSFRSLAPSASIRCTSCCPVP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 14B.

Table 14B. Comparison of NOV14a against NOV14b through NOV14e.		
Protein Sequence	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV14b	1..568 1..616	519/616 (84%) 524/616 (84%)
NOV14c	1..568 1..616	519/616 (84%) 524/616 (84%)
NOV14d	1..568 1..522	483/570 (84%) 485/570 (84%)
NOV14e	1..480 1..480	440/482 (91%) 443/482 (91%)

Further analysis of the NOV14a protein yielded the following properties shown in Table 14C.

5

Table 14C. Protein Sequence Properties NOV14a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 38 and 39

A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 14D.

Table 14D. Geneseq Results for NOV14a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB23625	Human secreted protein SEQ ID NO: 50 - Homo sapiens, 627 aa. [WO200049134-A1, 24-AUG-2000]	10..566 9..623	256/623 (41%) 386/623 (61%)	e-137
AAB36158	Novel human transporter protein SEQ ID NO: 2 - Homo sapiens, 627 aa. [WO200065055-A2, 02-NOV-2000]	10..566 9..623	256/623 (41%) 386/623 (61%)	e-137
AAB42213	Human ORFX ORF1977 polypeptide sequence SEQ ID NO:3954 - Homo sapiens, 627 aa. [WO200058473-A2, 05-OCT-2000]	10..566 9..623	256/623 (41%) 386/623 (61%)	e-136
AAB36164	Novel human transporter protein SEQ ID NO: 14 - Homo sapiens, 626 aa. [WO200065055-A2, 02-NOV-2000]	10..566 9..622	252/623 (40%) 382/623 (60%)	e-136
AAB36159	Novel human transporter protein SEQ ID NO: 4 - Homo sapiens, 627 aa. [WO200065055-A2, 02-NOV-2000]	10..566 9..623	256/623 (41%) 385/623 (61%)	e-136

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14E.

Table 14E. Public BLASTP Results for NOV14a				
Protein Accession Number	Protein/Organism/Length	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O57661	INTESTINAL SODIUM/LITHIUM-DEPENDENT DICARBOXYLATE TRANSPORTER (NA(+)/DICARBOXYLATE COTRANSPORTER) - Xenopus laevis (African clawed frog), 622 aa.	1..564 1..619	336/619 (54%) 444/619 (71%)	0.0
Q9ES88	NA/DICARBOXYLATE COTRANSPORTER (SOLUTE CARRIER FAMILY 13 (SODIUM-	1..561 1..567	311/572 (54%) 421/572 (73%)	e-179

	TRANSPORTER), MEMBER 2) - Mus musculus (Mouse), 586 aa.			
O35055	SODIUM/DICARBOXYLATE COTRANSPORTER 1 (NA(+)/DICARBOXYLATE COTRANSPORTER 1) (KIDNEY DICARBOXYLATE TRANSPORTER) (SDCT1) (ORGANIC ANION TRANSPORTER 1) (OAT1) - Rattus norvegicus (Rat), 587 aa.	1..562 1..568	311/572 (54%) 419/572 (72%)	e-179
Q13183	Renal sodium/dicarboxylate cotransporter (Na(+)/dicarboxylate cotransporter) - Homo sapiens (Human), 592 aa.	1..561 1..572	318/581 (54%) 428/581 (72%)	e-179
Q28615	Renal sodium/dicarboxylate cotransporter (Na(+)/dicarboxylate cotransporter) - Oryctolagus cuniculus (Rabbit), 593 aa.	1..562 1..576	300/586 (51%) 418/586 (71%)	e-172

PFam analysis predicts that the NOV14a protein contains the domains shown in the Table 14F.

Table 14F. Domain Analysis of NOV14a			
Pfam Domain	NOV14a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Na_sulph_symp: domain 1 of 1	6..554	163/604 (27%) 424/604 (70%)	8.3e-140

Example 15.

5   The NOV15 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis		
	SEQ ID NO: 53	1547 bp
NOV15a, CG57732-01 DNA Sequence	AACCCCCTTGACTGAAGCAATGGAGGGGGGTCCAGCTGTCTGCTGCCAGGATCCTCGG GCAGAGCTGGTAGAACGGGTGGCAGCCATCGATGTGACTCACTTGGAGGAGGCAGATG GTGGCCCAGAGCCTACTAGAAACGGTGTGGACCCCCCACCACGGGCCAGAGCTGCCTC TGTGATCCCTGGCAGTACTTCAAGACTGCTCCCAGCCCGGCCTAGCCTCTCAGCCAGG AAGCTTTCCCTACAGGAGCGGCCAGCAGGAAGCTATCTGGAGGCGCAGGCTGGGCCTT ATGCCACGGGGCCTGCCAGCCACATCTCCCCCCGGGCCTGGCGGAGGCCCACCATCGA GTCCCACCACGTGGCCATCTCAGATGCAGAGGACTGCGTGCAGCTGAACCAGTACAAG CTGCAGAGTGAGATTGGCAAGGGTGCCTACGGTGTGGTGAGGCTGGCCTACAACGAAA GTGAAGACAGACACTATGCAATGAAAGTCCTTTCCAAAAAGAAGTTACTGAAGCAGTA	

142

	TGGCTTTCCACGTCGCCCTCCCCCGAGAGGGTCCCAGGCTGCCCAGGGAGGACCAGCC AAGCAGCTGCTGCCCCTGGAGCGGGTGTACCAGGAGATTGCCATCCTGAAGAAGCTGG ACCACGTGAATGTGGTCAAACTGATCGAGGTACTGGATGACCCAGCTGAGGACAACCT CTATTTGCCCCGCATCCTTCTCCATAGGCCCGTCATGGAAGTGCCCTGTGACAAGCCC TTCTCGGAGGAGCAAGCTCGCCTCTACCTGCGGGACGTCATCCTGGGCCTCGAGTACG TGCACTGCCAGAAGATCGTCCACAGGGACATCAAGCCATCCAACCTGCTCCTGGGGGA TGATGGGCACGTGAAGATCGCCGACTTTGGCGTCAGCAACCAGTTTGAGGGGAACGAC GCTCAGCTGTCCAGCACGGCGGGAACCCCAGCATTCATGGCCCCCGAGGCCATTTCTG ATTCCGGCCAGAGCTTCAGTGGGAAGTTGGATGTATGGGCCACTGGCGTCACGTTGTA CTGCTTTGTCTATGGGAAGTGCCCATTCATCGACGATTTCATCCTGGCCCTCCACAGG AAGATCAAGAATGAGCCCGTGGTGTTTCCTGAGGAGCCAGAAATCAGCGAGGAGCTCA AGGACCTGATCCTGAAGATGTTAGACAAGAATCCCGAGACGAGAATTGGGGTGCCAGA CATCAAGTTGCACCCTTGGGTGACCAAGAACGGGGAGGAGCCCCTTCCTTCGGAGGAG GAGCACTGCAGCGTGGTGGAGGTGACAGAGGAGGAGGTTAAGAACTCAGTCAGGCTCA TCCCCAGCTGGACCACGGTGATCCTGGTGAAGTCCATGCTGAGGAAGCGTTCCTTTGG GAACCCGTTTGAGCCCCAAGCACGGAGGGAAGAGCGATCCATGTCTGCTCCAGGAAAC CTACTGGTGAAAGAAGGGTTTGGTGAAGGGGGCAAGAGCCCAGAGCTCCCCGGCGTCC AGGAAGACGAGGCTGCATCC**TGA**GCCCCTGCATGCACCC
	ORF Start: ATG at 20 / **ORF Stop: TGA at 1529**
	SEQ ID NO: 54 / **503 aa** / **MW at 55606.7kD**
NOV15a, CG57732-01 Protein Sequence	MEGGPAVCCQDPRAELVERVAAIDVTHLEEADGGPEPTRNGVDPPPRARAASVIPGST SRLLPARPSLSARKLSLQERPAGSYLEAQAGPYATGPASHISPRAWRRPTIESHHVAI SDAEDCVQLNQYKLQSEIGKGAYGVVRLAYNESEDRHYAMKVLSKKKLLKQYGFPRRP PPRGSQAAQGGPAKQLLPLERVYQEIAILKKLDHVNVVKLIEVLDDPAEDNLYLPRIL LHRPVMEVPCDKPFSEEQARLYLRDVILGLEYVHCQKIVHRDIKPSNLLLGDDGHVKI ADFGVSNQFEGNDAQLSSTAGTPAFMAPEAISDSGQSFSGKLDVWATGVTLYCFVYGK CPFIDDFILALHRKIKNEPVVFPEEPEISEELKDLILKMLDKNPETRIGVPDIKLHPW VTKNGEEPLPSEEEHCSVVEVTEEEVKNSVRLIPSWTTVILVKSMLRKRSFGNPFEPQ ARREERSMSAPGNLLVKEGFGEGGKSPELPGVQEDEAAS
	SEQ ID NO: 55 / **1611 bp**
NOV15b, CG57732-02 DNA Sequence	GCGCCCAGGTTCCCAACAAGGCTACGCAGAAGAACCCCCTTGACTGAAGC**ATG**GAGG GGGGTCCAGCTGTCTGCTGCCAGGATCCTCGGGCAGAGCTGGTAGAACGGGTGGCAGC CATCGATGTGACTCACTTGGAGGAGGCAGATGGTGGCCCAGAGCCTACTAGAAACGGT GTGGACCCCCCACCACGGGCCAGAGCTGCCTCTGTGATCCCTGGCAGTACTTCAAGAC TGCTCCCAGCCCGGCCTAGCCTCTCAGCCAGGAAGCTTTCCCTACAGGAGCGGCCAGC AGGAAGCTATCTGGAGGCGCAGGCTGGGCCTTATGCCACGGGGCCTGCCAGCCACATC TCCCCCCGGGCCTGGCGGAGGCCCACCATCGAGTCCCACCACGTGGCCATCTCAGATG CAGAGGACTGCGTGCAGCTGAACCAGTACAAGCTGCAGAGTGAGATTGGCAAGGGTGC CTACGGTGTGGTGAGGCTGGCCTACAACGAAAGTGAAGACAGACACTATGCAATGAAA GTCCTTTCCAAAAAGAAGTTACTGAAGCAGTATGCCTTTCCACGTCGCCCTCCCCCGA GAGGGTCCCAGGCTGCCCAGGGAGGACCAGCCAAGCAGCTGCTGCCCCTGGAGCGGGT GTACCAGGAGATTGCCATCCTGAAGAAGCTGGACCACGTGAATGTGGTCAAACTGATC GAGGTCCTGGATGACCCAGCTGAGGACAACCTCTATTTGGTGTTTGACCTCCTGAGAA AGGGGCCCGTCATGGAAGTGCCCTGTGACAAGTCCTTCTCGGAGGAGCAAGCTCGCCT CTACCTGCGGGACGTCATCCTGGGCCTCGAGTACTTGCACTGCCAGAAGATCGTCCAC AGGGACATCAAGCCATCCAACCTGCTCCTGGGGGATGATGGGCACGTGAAGATCGCCG ACTTTGGCGTCAGCAACCAGTTTGAGGGGAACGACGCTCAGCTGTCCAGCACGGCGGG AACCCCAGCATTCATGGCCCCCGAGGCCATTTCTGATTCCGGCCAGAGCTTCAGTGGG AAGGCCTTGGATGTATGGGCCACTGGCGTCACTGCTTTGTCTATGGGAAGT GCCCGTTCATCGACGATTTCATCCTGGCCCTCCACAGGAAGATCAAGAATGAGCCCGT GGTGTTTCCTGAGGGGCCAGAAATCAGCGAGGAGCTCAAGGACCTGATCCTGAAGATG TTAGACAAGAATCCCGAGACGAGAATTGGGGTGCCAGACATCAAGTTGCACCCTTGGG TGACCAAGAACGGGGAGGAGCCCCTTCCTTCGGAGGAGGAGCACTGCAGCGTGGTGGA GGTGACAGAGGAGGAGGTTAAGAACTCAGTCAGGCTCATCCCCAGCTGGACCACGGTG ATCCTGGTGAAGTCCATGCTGAGGAAGCGTTCCTTTGGGAACCCGTTTGAGCCCCAAG CACGGAGGGAAGAGCGATCCATGTCTGCTCCAGGAAACCTACTGGTGAAAGAAGGGTT TGGTGAAGGGGGCAAGAGCCCAGAGCTCCCCGGCGTCCAGGAAGACGAGGCTGCATCC **TGA**GCCCCTGCATGCACCCAGGGCCACCCGGCAGCACACTCATCC
	ORF Start: ATG at 52 / **ORF Stop: TGA at 1567**
	SEQ ID NO: 56 / **505 aa** / **MW at 55652.7kD**
NOV15b, CG57732-02 Protein Sequence	MEGGPAVCCQDPRAELVERVAAIDVTHLEEADGGPEPTRNGVDPPPRARAASVIPGST SRLLPARPSLSARKLSLQERPAGSYLEAQAGPYATGPASHISPRAWRRPTIESHHVAI SDAEDCVQLNQYKLQSEIGKGAYGVVRLAYNESEDRHYAMKVLSKKKLLKQYGFPRRP PPRGSQAAQGGPAKQLLPLERVYQEIAILKKLDHVNVVKLIEVLDDPAEDNLYLVFDL LRKGPVMEVPCDKSFSEEQARLYLRDVILGLEYLHCQKIVHRDIKPSNLLLGDDGHVK IADFGVSNQFEGNDAQLSSTAGTPAFMAPEAISDSGQSFSGKALDVWATGVTLYCFVY GKCPFIDDFILALHRKIKNEPVVFPEGPEISEELKDLILKMLDKNPETRIGVPDIKLH PWVTKNGEEPLPSEEEHCSVVEVTEEEVKNSVRLIPSWTTVILVKSMLRKRSFGNPFE PQARREERSMSAPGNLLVKEGFGEGGKSPELPGVQEDEAAS

	SEQ ID NO: 57	1725 bp
NOV15c, CG57732-03 DNA Sequence	GCGCCCAGGTTCCCAACAAGGCTACGCAGAAGAACCCCCTTGACTGAAGTAATGGAGG GGGGTCCAGCTGTCTGCTGCCAGGATCCTCGGGCAGAGCTGGTAGAACGGGTGGCAGC CATCGATGTGACTCACTTGGAGGAGGCAGATGGTGGCCCAGAGCCTACTAGAAACGGT GTGGACCCCCCACCACGGGCCAGAGCTGCCTCTGTGATCCCTGGCAGTACTTCAAGAC TGCTCCCAGCCCGGCCTAGCCTCTCAGCCAGGAAGCTTTCCCTACAGGAGCGGCCAGC AGGAAGCTATCTGGAGGCGCAGGCTGGGCCTTATGCCACGGGGCCTGCCAGCCACATC TCCCCCCGGGCCTGGCGGAGGCCCACCATCGAGTCCCACCACGTGGCCATCTCAGATG CAGAGGACTGCGTGCAGCTGAACCAGTACAAGCTGCAGAGTGAGATTGGCAAGGGTGC CTACGGTGTGGTGAGGCTGGCCTACAACGAAAGTGAAGACAGACACTATGCAATGAAA GTCCTTTCCAAAAAGAAGTTACTGAAGCAGTATGGCTTTCCACGTCGCCCTCCCCCGA GAGGGTCCCAGGCTGCCCAGGGAGGACCAGCCAAGCAGCTGCTGCCCCTGGAGCGGGT GTACCAGGAGATTGCCATCCTGAAGAAGCTGGACCACGTGAATGTGGTCAAACTGATC GAGGTCCTGGATGACCCGGCTGAGGACAACCTCTATTTGGCCCTGCAGAACCAGGCCC AGAATATCCAGTTAGATTCAACAAATATCGCCAAGTCCCACTCCCTGCTTCCCTCTGA GCAGCAAGACAGTGGATCCACGTGGGCTGCGCGCTCAGTGTTTGACCTCCTGAGAAAG GGGCCCGTCATGGAAGTGCCCTGTGACAAGCCCTTCTCGGAGGAGCAAGCTCGCCTCT ACCTGCGGGACGTCATCCTGGGCCTCGAGTACTTGCACTGCCAGAAGATCGTCCACAG GGACATCAAGCCATCCAACCTGCTCCTGGGGGATGATGGGCACGTGAAGATCGCCGAC TTTGGCGTCAGCAACCAGTTTGAGGGGAACGACGCTCAGCTGTCCAGCACGGCGGGAA CCCCAGCATTCATGGCCCCCGAGGCCATTTCTGATTCCGGCCAGAGCTTCAGTGGGAA GGCCTTGGATGTATGGGCCACTGGCGTCACGTTGTACTGCTTTGTCTATGGGAAGTGC CCGTTCATCGACGATTTCATCCTGGCCCTCCACAGGAAGACCAAGAATGAGCCCGTGG TGTTTCCTGAGGGGCCAGAAATCAGCGAGGAGCTCAAGGACCTGATCCTGAAGATGTT AGACAAGAATCCCGAGACGAGAATTGGGGTGCCAGACATCAAGTTGCACCCTTGGGTG ACCAAGAACGGGGAGGAGCCCCCTTCCTTCGGAGGAGGAGCACTGCAGCGTGGTGGAGG TGACAGAGGAGGAGGTTAAGAACTCAGTCAGGCTCATCCCCAGCTGGACCACGGTGAT CCTGGTGAAGTCCATGCTGAGGAAGCGTTCCTTTGGGAACCCGTTTGAGCCCCAAGCA CGGAGGGAAGAGCGATCCATGTCTGCTCCAGGAAACCTACTGGTGAAAGAAGGGTTTG GTGAAGGGGGCAAGAGCCCAGAGCTCCCCGGCGTCCAGGAAGACGAGGCTGCATCCTG AGCCCCTGCATGCACCCAGGGCCACCCGGCAGCACACTCATCC
	ORF Start: ATG at 52	ORF Stop: TGA at 1681
	SEQ ID NO: 58	543 aa	MW at 59729.0kD
NOV15c, CG57732-03 Protein Sequence	MEGGPAVCCQDPRAELVERVAAIDVTHLEEADGGPEPTRNGVDPPPRARAASVIPGST SRLLPARPSLSARKLSLQERPAGSYLEAQAGPYATGPASHISPRAWRRPTIESHHVAI SDAEDCVQLNQYKLQSEIGKGAYGVVRLAYNESEDRHYAMKVLSKKKLLKQYGFPRRP PPRGSQAAQGGPAKQLLPLERVYQEIAILKKLDHVNVVKLIEVLDDPAEDNLYLALQN QAQNIQLDSTNIAKSHSLLPSEQQDSGSTWAARSVFDLLRKGPVMEVPCDKPFSEEQA RLYLRDVILGLEYLHCQKIVHRDIKPSNLLLGDDGHVKIADFGVSNQFEGNDAQLSST AGTPAFMAPEAISDSGQSFSGKALDVWATGVTLYCFVYGKCPFIDDFILALHRKTKNE PVVFPEGPEISEELKDLILKMLDKNPETRIGVPDIKLHPWVTKNGEEPLPSEEEHCSV VEVTEEEVKNSVRLIPSWTTVILVKSMLRKRSFGNPFEPQARREERSMSAPGNLLVKE GFGEGGKSPELPGVQEDEAAS

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 15B.

Table 15B. Comparison of NOV15a against NOV15b through NOV15c.		
Protein Sequence	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV15b	1..503 1..505	495/505 (98%) 497/505 (98%)
NOV15c	1..503 1..543	492/543 (90%) 495/543 (90%)

Further analysis of the NOV15a protein yielded the following properties shown in Table 15C.

144

Table 15C. Protein Sequence Properties NOV15a	
PSort analysis:	0.7600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 15D.

Table 15D. Geneseq Results for NOV15a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU03510	Human protein kinase #10 - Homo sapiens, 513 aa. [WO200138503-A2, 31-MAY-2001]	1..503 1..513	496/513 (96%) 498/513 (96%)	0.0
AAE04361	Human kinase (PKIN)-2 - Homo sapiens, 513 aa. [WO200146397-A2, 28-JUN-2001]	1..503 1..513	496/513 (96%) 498/513 (96%)	0.0
AAY44239	Human cell signalling protein-2 - Homo sapiens, 540 aa. [WO9958558-A2, 18-NOV-1999]	64..500 90..538	289/450 (64%) 367/450 (81%)	e-165
AAM40450	Human polypeptide SEQ ID NO 5381 - Homo sapiens, 680 aa. [WO200153312-A1, 26-JUL-2001]	64..482 128..558	283/432 (65%) 356/432 (81%)	e-162
AAM40449	Human polypeptide SEQ ID NO 5380 - Homo sapiens, 680 aa. [WO200153312-A1, 26-JUL-2001]	64..482 128..558	283/432 (65%) 356/432 (81%)	e-162

5   In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15E.

10

Table 15E. Public BLASTP Results for NOV15a				
Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BQH3	HYPOTHETICAL 55.7 KDA PROTEIN - Homo sapiens (Human), 505 aa.	1..503 1..505	497/505 (98%) 499/505 (98%)	0.0
P97756	CA2+/CALMODULIN-DEPENDENT PROTEIN KINASE IV KINASE ISOFORM - Rattus norvegicus (Rat), 505 aa.	1..503 1..505	465/505 (92%) 478/505 (94%)	0.0
AAH17529	SIMILAR TO CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE KINASE 1, ALPHA - Mus musculus (Mouse), 505 aa.	1..503 1..505	464/505 (91%) 478/505 (93%)	0.0
Q64572	CA2+/CALMODULIN-DEPENDENT PROTEIN KINASE KINASE (EC 2.7.1.37) - Rattus norvegicus (Rat), 505 aa.	1..503 1..505	463/505 (91%) 476/505 (93%)	0.0
Q9R054	CALCIUM/CALMODULIN DEPENDENT PROTEIN KINASE KINASE ALPHA - Mus musculus (Mouse), 505 aa.	1..503 1..505	454/505 (89%) 471/505 (92%)	0.0

PFam analysis predicts that the NOV15a protein contains the domains shown in the Table 15F.

Table 15F. Domain Analysis of NOV15a			
Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Pkinase: domain 1 of 2	128..228	28/101 (28%) 81/101 (80%)	8.4e-16
Pkinase: domain 2 of 2	245..407	70/201 (35%) 129/201 (64%)	1.7e-52

Example 16.

The NOV16 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 16A.

146

Table 16A. NOV16 Sequence Analysis		
	SEQ ID NO: 59	688 bp
NOV16a, CG57709-01 DNA Sequence	GACGCGGACCCGCCATGGCGCGGAAGAAGGTGCGTCCGCGGCTGATCGCGGAGCTGGC CCGCCGCGTGCGCGCCCTGCGGGAGCAACTGAACAGGCCGCGCGACTCCCAGCTCTAC GCGGTGGACTACGAGACCTTGACGCGGCCGTTCTCTGGACGCCGGCTGCCGGTCCGGG CCTGGGCCGACGTGCGCCGCGAGAGCCGCCTCTTGCAGCTGCTCGGCCGCCTCCCGCT CTTCGGCCTGGGCCGCCTGGTCACGCGCAAGTCCTGGCTGTGGCAGCACGACGAGCCG TGCTACTGGCGCCTCACGCGGGTGCGGCCCGACTACACGGCGCAGAACTTGGACCACG GGAAGGCCTGGGGCATCCTGACCTTCAAAGGTAAGGCTCGGGAGAGCGCGCGGGAGAT CGAACACGTCATGTACCATGACTGGCGGCTGGTGCCCAAGCACGAGGAGGAGGCCTTC ACCGCGTTCACGCCGGCGCCGGAAGACAGCCTGGCCTCCGTGCCGTACCCGCCTCTCC TCCGGGCCATGATTATCGCAGAACGACAGAAAAATGGAGACACAAGCACCGAGGAGCC CATGCTGAATGTGCAGAGGATACGCATGGAACCCTGGGATTACCCTGCAAAACAGGAA GACAAAGGAAGGGCCAAGGGCACCCCCGTCTAGAATGCCAGAACCAGCGG
	ORF Start: ATG at 15	ORF Stop: TAG at 669
	SEQ ID NO: 60	218 aa	MW at 25647.2kD
NOV16a, CG57709-01 Protein Sequence	MARKKVRPRLIAELARRVRALREQLNRPRDSQLYAVDYETLTRPFSGRRLPVRAWADV RRESRLLQLLGRLPLFGLGRLVTRKSWLWQHDEPCYWRLTRVRPDYTAQNLDHGKAWG ILTFKGKARESAREIEHVMYHDWRLVPKHEEEAFTAFTPAPEDSLASVPYPPLLRAMI IAERQKNGDTSTEEPMLNVQRIRMEPWDYPAKQEDKGRAKGTPV	

Further analysis of the NOV16a protein yielded the following properties shown in Table 16B.

Table 16B. Protein Sequence Properties NOV16a	
PSort analysis:	0.9081 probability located in mitochondrial matrix space; 0.6000 probability located in mitochondrial inner membrane; 0.6000 probability located in mitochondrial intermembrane space; 0.6000 probability located in mitochondrial outer membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 16C.

Table 16C. Geneseq Results for NOV16a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG81356	Human AFP protein sequence SEQ ID NO:230 - Homo sapiens, 218 aa. [WO200129221-A2, 26-APR-2001]	1..218 1..218	212/218 (97%) 212/218 (97%)	e-125
AAU30525	Novel human secreted protein	135..218 1..84	84/84 (100%) 84/84 (100%)	3e-45

147

	[WO200179449-A2, 25-OCT-2001]			
AAU30526	Novel human secreted protein #1017 - Homo sapiens, 62 aa. [WO200179449-A2, 25-OCT-2001]	187..217 12..42	31/31 (100%) 31/31 (100%)	4e-12

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16D.

Table 16D. Public BLASTP Results for NOV16a				
Protein Accession Number	Protein/Organism/Length	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BVI7	HYPOTHETICAL 25.7 KDA PROTEIN - Homo sapiens (Human), 218 aa.	1..218 1..218	214/218 (98%) 214/218 (98%)	e-125
P82930	MITOCHONDRIAL 28S RIBOSOMAL PROTEIN S34 (MRP-S34) – Homo sapiens (Human), 218 aa.	1..218 1..218	213/218 (97%) 213/218 (97%)	e-124
CAC38606	SEQUENCE 229 FROM PATENT WO0129221 - Homo sapiens (Human), 218 aa.	1..218 1..218	212/218 (97%) 212/218 (97%)	e-124
Q9JIK9	TCE2 (0610007F04RIK PROTEIN) - Mus musculus (Mouse), 218 aa.	1..218 1..218	194/218 (88%) 205/218 (93%)	e-114
Q9D957	0610007F04RIK PROTEIN - Mus musculus (Mouse), 218 aa.	1..218 1..218	193/218 (88%) 205/218 (93%)	e-114

PFam analysis predicts that the NOV16a protein contains the domains shown in the Table 16E.

Table 16E. Domain Analysis of NOV16a			
Pfam Domain	NOV16a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

148

Example 17.

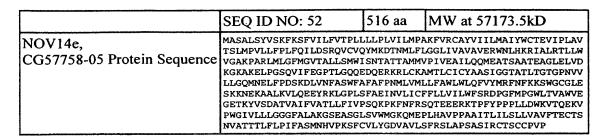The NOV17 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 17A.

<table>
<tr><td colspan="4" align="center">Table 17A. NOV17 Sequence Analysis</td></tr>
<tr><td></td><td>SEQ ID NO: 61</td><td colspan="2">894 bp</td></tr>
<tr><td>NOV17a,
CG57700-01 DNA Sequence</td><td colspan="3">CTCCGTGACCATGAAGGTCAAGGTCATCCCCGTGCTCGAGGACAACTACATGTACCTG
GTCATCGAGGAGCTCACGCGCGAGGCGGTGGCCGTGGACGTGGCTGTGCCCAAGAGGC
TGCTGGAGATCGTGGGCCGGGAGGGGGTGTCTCTGACCGCTGTGCTGACCACCCACCA
TCACTGGGACCACGCGCGGGGAAACCCGGAGCTGGCGCGGCTTCGTCCCGGGCTGGCG
GTGCTGGGCGCGGACGAGCGCATCTTCTCGCTGACGCGCAGGCTGGCGCACGGCGAGG
AGCTGCAGTTCGGGGCCATCCACGTGCGTTGCCTCCTGACGCCCGGCCACACCGCCGG
CCACATGAGCTACTTCCTGTGGGAGGACGATTGCCCGGACCCACCCGCCCTGTTCTCG
GGTGGCGACGCGCTGTCGGTGGCCGGCTGCGGCTCGTGCCTGGAGGGCAGCGCCCAGC
AGATGTACCAGAGCCTGGCCGAGCTGGGTACCCTGCCCCCCGAGACGAAGGTGTTCTG
CGGCCACGAGCACACGCTTAGCAACCTGGAGTTTGCCCAGAAAGTGGAGCCCTGCAAC
GACCACGTGAGAGCCAAGCTGTCCTGGGCTCAGAAGAGGGATGAGGATGACGTGCCCA
CTGTGCCGTCGACTCTGGGCGAGGAGCGCCTCTACAACCCCTTCCTGCGGGTGGCGGA
GGAGCCGGTGCGCAAGTTCACGGGCAAGGCGGTCCCCGCCGACGTCCTGGAGGCGCTA
TGCAAGGAGCGGGCGCGCTTCGAACAGGCGGGCGAGCCGCGGCAGCCACAGGCGCGGG
CCCTCCTTGCGCTGCAGTGGGGGCTCCTGAGTGCAGCCCCACACGACTGAGCCACCCA
GACCCTCACAGGGCTGGGGCCTGC</td></tr>
<tr><td></td><td>ORF Start: ATG at 11</td><td colspan="2">ORF Stop: TGA at 860</td></tr>
<tr><td></td><td>SEQ ID NO: 62</td><td>283 aa</td><td>MW at 31262.3kD</td></tr>
<tr><td>NOV17a,
CG57700-01 Protein Sequence</td><td colspan="3">MKVKVIPVLEDNYMYLVIEELTREAVAVDVAVPKRLLEIVGREGVSLTAVLTTHHHWD
HARGNPELARLRPGLAVLGADERIFSLTRRLAHGEELQFGAIHVRCLLTPGHTAGHMS
YFLWEDDCPDPPALFSGGDALSVAGCGSCLEGSAQQMYQSLAELGTLPPETKVFCGHE
HTLSNLEFAQKVEPCNDHVRAKLSWAQKRDEDDVPTVPSTLGEERLYNPFLRVAEEPV
RKFTGKAVPADVLEALCKERARFEQAGEPRQPQARALLALQWGLLSAAPHD</td></tr>
<tr><td></td><td>SEQ ID NO: 63</td><td colspan="2">888 bp</td></tr>
<tr><td>NOV17b,
CG57700-02 DNA Sequence</td><td colspan="3">CTCCGTGACCATGAAGGTCAAGGTCATCCCCGTGCTCGAGGACAACTACATGTACCTG
GTCATCGAGGAGCTCACGCGCGAGGCGGTGGCCGTGGACGTGGCTGTGCCCAAGAGGC
TGCTGGAGATCGTGGGCCGGGAGGGGGTGTCTCTGACCGCTGTGCTGACCACCCACCA
TCACTGGGACCACGCGCGGGGAAACCCGGAGCTGGCGCGGCTTCGTCCCGGGCTGGCG
GTGCTGGGCGCGGACGAGCGCATCTTCTCGCTGACGCGCAGGCTGGCGCACGGCGAGG
AGCTGCAGTTCGGGGCCATCCACGTGCGTTGCCTCCTGACGCCCGGCCACACCGCCGG
CCACATGAGCTACTTCCTGTGGGAGGACGATTGCCCGGACCCACCCGCCCTGTTCTCG
GGCGACGCGCTGTCGGTGGCCGGCTGCGGCTCGTGCCTGGAGGGCAGCGCCCAGCAGA
TGTACCAGAGCCTGGCCGAGCTGGGTACCCTGCCCCCCGAGACGAAGGTGTTCTGCGG
CCACGAGCACACACTTAGCAACCTGGAGTTTGCCCAGAAAGTGGAGCCCTGCAACGAC
CACGTGAGAGCCAAGCTGTCCTGGGCTAAGAAGAGGGATGAGGATGACGTGCCCACTG
TGCCGTCGACTCTGGGCGAGGAGCGCCTCTACAACCCCTTCCTGCGGGTGGCAGAGGA
GCCGGTGCGCAAGTTCACGGGCAAGGCGGTCCCCGCCGACGTCCTGGAGGCGCTATGC
AAGGAGCGGGCGCGCTTCGAACAGGCGGGCGAGCCGCGGCAGCCACAGGCGCGGGCCC
TCCTTGCGCTGCAGTGGGGGCTCCTGAGTGCAGCCCCACACGACTGAGCCACCCAGAC
CCTCACAGGGCTGGGCCT</td></tr>
<tr><td></td><td>ORF Start: ATG at 11</td><td colspan="2">ORF Stop: TGA at 857</td></tr>
<tr><td></td><td>SEQ ID NO: 64</td><td>282 aa</td><td>MW at 31205.3kD</td></tr>
<tr><td>NOV17b,
CG57700-02 Protein Sequence</td><td colspan="3">MKVKVIPVLEDNYMYLVIEELTREAVAVDVAVPKRLLEIVGREGVSLTAVLTTHHHWD
HARGNPELARLRPGLAVLGADERIFSLTRRLAHGEELQFGAIHVRCLLTPGHTAGHMS
YFLWEDDCPDPPALFSGDALSVAGCGSCLEGSAQQMYQSLAELGTLPPETKVFCGHEH
TLSNLEFAQKVEPCNDHVRAKLSWAKKRDEDDVPTVPSTLGEERLYNPFLRVAEEPVR
KFTGKAVPADVLEALCKERARFEQAGEPRQPQARALLALQWGLLSAAPHD</td></tr>
<tr><td></td><td>SEQ ID NO: 65</td><td colspan="2">882 bp</td></tr>
<tr><td>NOV17c,
CG57700-03 DNA Sequence</td><td colspan="3">ACCATGAAGGTCAAGGTCATCCCCGTGCTCGAGGACAACTACATGTACCTGGTCATCG
AGGAGCTCACGCGCGAGGCGGTGGCCGTGGACGTGGCTGTGCCCAAGAGGCTGCTGGA
GATCGTGGGCCGGGAGGGGGTGTCTCTGACCGCTGTGCTGACCACCCACCATCACTGG
GACCACGCGCGGGGAAACCCGGAGCTGGCGCGGCTTCGTCCCGGGCTGGCGGTGCTGG
GCGCGGACGAGCGCATCTTCTCGCTGACGCGCAGGCTGGCGCACGGCGAGGAGCTGCG
GTTCGGGGCCATCCACGTGCGTTGCCTCCTGACGCCCGGCCACACCGCCGGCCACATG
AGCTACTTCCTGTGGGAGGACGATTGCCCGGACCCACCCGCCCTGTTCTCGGGGCGACG
CGCTGTCGGTGGCCGGCTGCGGCTCGTGCCTGGAGGGCAGCGCCCAGCAGATGTACCA</td></tr>
</table>

	GAGCCTGGCCGAGCTGGGTACCCTGCCCCCCGAGACGAAGGTGTTCTGCGGCCACGAG CACACGCTTAGCAACCTGGAGTTTGCCCAGAAAGTGGAGCCCTGCAACGACCACGTGA GAGCCAAGCTGTCCTGGGCTAAGAAGAGGGATGAGGATGACGTGCCCACTGTGCCGTC GACTCTGGGCGAGGAGCGCCTCTACAACCCCTTCCTGCGGGTGGCAGAGGAGCCGGTG CGCAAGTTCACGGGCAAGGCGGTCCCCGCCGACGTCCTGGAGGCGCTATGCAAGGAGC GGGCGCGCTCCGAACAGGCGGGCGAGCCGCGGCAGCCACAGGCGCGGGCCCTCCTTGC GCTGCAGTGGGGGCTCCTGAGTGCAGCCCCACACGAC**TGA**GCCACCCAGACCCTCACA GGGCTGGGGCCT
	ORF Start: ATG at 4 ORF Stop: TGA at 850
	SEQ ID NO: 66 282 aa MW at 31173.2kD
NOV17c, CG57700-03 Protein Sequence	MKVKVIPVLEDNYMYLVIEELTREAVAVDVAVPKRLLEIVGREGVSLTAVLTTHHHWD HARGNPELARLRPGLAVLGADERIFSLTRRLAHGEELRFGAIHVRCLLTPGHTAGHMS YFLWEDDCPDPPALFSGDALSVAGCGSCLEGSAQQMYQSLAELGTLPPETKVFCGHEH TLSNLEFAQKVEPCNDHVRAKLSWAKKRDEDDVPTVPSTLGEERLYNPFLRVAEEPVR KFTGKAVPADVLEALCKERARSEQAGEPRQPQARALLALQWGLLSAAPHD
	SEQ ID NO: 67 855 bp
NOV17d, CG57700-04 DNA Sequence	<u>ACC</u>**ATG**AAGGTCAAGGTCATCCCCGTGCTCGAGGACAACTACATGTACCTGGTCATCG AGGAGCTCACGCGCGAGGCGGTGGCCGTGGACGTGGCTGTGCCCAAGAGGCTGCTGGA GATCGTGGGCCGGGAGGGGGTGTCTCTGACCGCTGTGCTGACCACCCACTATCACTGG GACCACGCGCGGGGAAACCCGGAGCTGGCGCGGCTTCGTCCCGGGCTGGCGGTGCTGG GCGCGGACGAGCGCATCTTCTCGCTGACGCGCAGGCTGGCGCACGGCGAGGAGCTGCG GTTCGGGGCCATCCACGTGCGTTGCCTCCTGACGCCCGGCCACACCGCCGGCCACATG AGCTACTTCCTGTGGGAGGACGATTGCCCGGACCCACCCGCCCTGTTCTCGGGCGACG CGCTGTCGGTGGCCGGCTGCGGCTCGTGCCTGGAGGGCAGCGCCCAGCAGATGTACCA GAGCCTGGCCGAGCTGGGTACCCTGCCCCCCGAGACGAAGGTGTTCTGCGGCCACGAG CACACGCTTAGCAACCTGGAGTTTGCCCAGAAAGTGGAGCCCTGCAACGACCACAAGA GGGATGAGGATGACGTGCCCACTGTGCCGTCGACTCTGGGCGAGGAGCGCCTCTACAA CCCCTTCCTGCGGGTGGCAGAGGAGCCGGTGCAAGTTCACGGGCAAGGCGGTCCCC GCCGACGTCCTGGAGGCGCTATGCAAGGAGCGGGCGCGCTTCGAACAGGCGGGCGAGC CGCGGCAGCCACAGGCGCGGGCCCTCCTTGCGCTGCAGTGGGGGCTCCTGAGTGCAGC CCCACACGAC**TGA**GCCACCCAGACCCTCACAGGGCTGGGGCCT
	ORF Start: ATG at 4 ORF Stop: TGA at 823
	SEQ ID NO: 68 273 aa MW at 30219.1kD
NOV17d, CG57700-04 Protein Sequence	MKVKVIPVLEDNYMYLVIEELTREAVAVDVAVPKRLLEIVGREGVSLTAVLTTHYHWD HARGNPELARLRPGLAVLGADERIFSLTRRLAHGEELRFGAIHVRCLLTPGHTAGHMS YFLWEDDCPDPPALFSGDALSVAGCGSCLEGSAQQMYQSLAELGTLPPETKVFCGHEH TLSNLEFAQKVEPCNDHKRDEDDVPTVPSTLGEERLYNPFLRVAEEPVRKFTGKAVPA DVLEALCKERARFEQAGEPRQPQARALLALQWGLLSAAPHD

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 17B.

Table 17B. Comparison of NOV17a against NOV17b through NOV17d.		
Protein Sequence	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV17b	1..283 1..282	281/283 (99%) 282/283 (99%)
NOV17c	1..283 1..282	279/283 (98%) 281/283 (98%)
NOV17d	1..283 1..273	271/283 (95%) 273/283 (95%)

Further analysis of the NOV17a protein yielded the following properties shown in Table 17C.

Table 17C. Protein Sequence Properties NOV17a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1682 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 17D.

Table 17D. Geneseq Results for NOV17a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW80783	Human bisphosphonate binding protein, DP1 (hDP1) - Homo sapiens, 260 aa. [WO9836064-A1, 20-AUG-1998]	1..256 1..256	128/257 (49%) 184/257 (70%)	6e-72
AAG10987	Arabidopsis thaliana protein fragment SEQ ID NO: 9531 - Arabidopsis thaliana, 258 aa. [EP1033405-A2, 06-SEP-2000]	1..245 1..246	107/248 (43%) 160/248 (64%)	5e-53
AAG10986	Arabidopsis thaliana protein fragment SEQ ID NO: 9530 - Arabidopsis thaliana, 268 aa. [EP1033405-A2, 06-SEP-2000]	1..245 11..256	107/248 (43%) 160/248 (64%)	5e-53
AAM78721	Human protein SEQ ID NO 1383 - Homo sapiens, 385 aa. [WO200157190-A2, 09-AUG-2001]	1..226 119..344	100/227 (44%) 135/227 (59%)	6e-45
AAY71110	Human Hydrolase protein-8 (HYDRL-8) - Homo sapiens, 361 aa. [WO200028045-A2, 18-MAY-2000]	1..226 95..320	100/227 (44%) 135/227 (59%)	6e-45

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17E.

151

Table 17E. Public BLASTP Results for NOV17a				
Protein Accession Number	Protein/Organism/Length	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BT45	SIMILAR TO RIKEN CDNA 1500017E18 GENE - Homo sapiens (Human), 282 aa.	1..283 1..282	280/283 (98%) 282/283 (98%)	e-163
Q9DB32	1500017E18RIK PROTEIN - Mus musculus (Mouse), 283 aa.	1..278 1..278	231/279 (82%) 251/279 (89%)	e-133
Q96S11	SIMILAR TO HAGH - Homo sapiens (Human), 218 aa.	1..228 1..218	217/228 (95%) 218/228 (95%)	e-123
Q96NR5	CDNA FLJ30279 FIS, CLONE BRACE2002772, MODERATELY SIMILAR TO HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6) - Homo sapiens (Human), 202 aa.	1..133 1..133	132/133 (99%) 133/133 (99%)	3e-73
O35952	Hydroxyacylglutathione hydrolase (EC 3.1.2.6) (Glyoxalase II) (Glx II) (Round spermatid protein RSP29) - Rattus norvegicus (Rat), 260 aa.	1..256 1..256	128/257 (49%) 184/257 (70%)	1e-71

PFam analysis predicts that the NOV17a protein contains the domains shown in the Table 17F.

Table 17F. Domain Analysis of NOV17a			
Pfam Domain	NOV17a Match Region	Identities/ Similarities for the Matched Region	Expect Value
lactamase_B: domain 1 of 1	7..173	55/221 (25%) 129/221 (58%)	5.8e-32

Example 18.

The NOV18 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 18A.

Table 18A. NOV18 Sequence Analysis		
	SEQ ID NO: 69	2109 bp
NOV18a, CG58553-01 DNA Sequence	GGGTCCGGCGGGCATCGGCAAGACCATGGCGGCCAAAAATATCCTGTACGACTGGGCG GCGGGCAAGCTGTACCAGGGCCAGGTGGACTTCGCCTTCTTCATGCCCTGCGGCGAGC TGCTGGAGAGGCCGGGCACGCGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGA CCGCGGCGCGCCGGTGCCGCAGATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTG GACGGCGCGGACGAGCTGCCGGCGCTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACC CCTTCGAGGCGGCGAGCGGCGCGCGGGTGCTAGGCGGGCTGCTGAGTAAGGCGCTGCT GCCCACGGCCCTCCTGCTGGTGACCACGCGCCGCCGCCGCCCCGGGAGGCTGCAGGGC CGCCTGTGTTCCCCGCAGTGCGCCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAAGA AGTATTTCTACAAGTTCTTCCGGGATGAGAGGAGGGCCGAGCGCGCCTACCGCTTCGT GAAGGAGAACGAGACGCTGTTCGCGCTGTGCTTCGTGCCCTTCGTGTGCTGGATCGTG TGCACCGTGCTGCGCCAGCAGCTGGAGCTCGGTCGGGACCTGTCGCGCACGTCCAAGA CCACCACGTCAGTGTACCTGCTTTTCATCACCAGCGTTCTGAGCTCGGCTCCGGTAGC CGACGGGCCCCGGTTGCAGGGCGACCTGCGCAATCTGTGCCGCCTGGCCCGCGAGGGC GTCCTCGGACGCAGGGCGCAGTTTGCCGAGAAGGAACTGGAGCAACTGGAGCTTCGTG GCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGGAGCTGCCGGGCGTGCTGGAGAC AGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGGAGTCCTTCGCGGCACTGTCC TACCTGCTGGAGGACGGCGGGGTGCCCAGGACCGCGGCTGGCGGCGTTGGGACACTCC TGCGTGGGGACGCCCAGCCGCACAGCCACTTGGTGCTCACCACGCGCTTCCTCTTCGG ACTGCTGAGCGCGGAGCGGATGCGCGACATCGAGCGCCACTTCGGCTGCATGGTTTCA GAGCGTGTGAAGCAGGAGCCCTGCGGTGGGTGCAGGGACAGGGACAGGGCTGCCCCG GAGTGGCACCAGAGGTGACCGAGGGGGCCAAAGGGCTCGAGGACACCGAAGAGCCAGA GGAGGAGGAGGAGGGAGAGGAGCCCAACTACCCACTGGAGTTGCTGTACTGCCTGTAC GAGACGCAGGAGGACGCGTTTGTGCGCCAAGCCCTGGGCCGGTTCCCGGAGCTGGCGC TGCAGCGAGTGCGCTTCTGCCGCCATGGACGTGGCTGTTCTGAGCTACTGCGTGAGGTG CTGCCCTGCTGCACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTGCGCAGGAG AAGAAGAAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCTGGGCACCACAAAACAAC TGCCAGCCTCCCTTCTTCATCCACTCTTTCAGGCAATGACTGACCCACTGTGCCATCT GAGCAGCCTCACGCTGTCCCACTGCAAACTCCCTGACGCGGTCTGCCGAGACCTTTCT GAGGCCCTGAGGGCAGCCCCCGCACTGACGGAGCTGGGCCTCCTCCACAACAGGCTCA GTGAGGCAGGACTGCGTATGCTGAGTGAGGGCCTAGCCTGGCCGCAGTGCAGGGTGCA GACGGTCAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTCCAGTACCTGGTGGGTATG CTTCGGCAGAGCCCTGCCCTGACCACCCTGGATCTCAGCGGCTGCCAACTGCCCCGCCC CCATGGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGGATGCGGCCTGCAGACCCT CAGTCTGGCCTCTGTGGAGCTGAGCGAGCAGTCACTACAGGAGCTTCAGGCTGTGAAG AGAGCAAAGCCGGATCTGGTCATCACACACCCAGCGCTGGACGGCCACCCACAACCTC CCAAGGAACTCATCTCGACCTTCTGAGGCTCTGGTGGCCAGAGCAGGGTGGAAGACCC TAGTCAAAGTCCCTGTGGAGA
	ORF Start: ATG at 26	ORF Stop: TGA at 2054
	SEQ ID NO: 70	676 aa	MW at 74650.3kD
NOV18a, CG58553-01 Protein Sequence	MAAKNILYDWAAGKLYQGQVDFAFFMPCGELLERPGTRSLADLILDQCPDRGAPVPQM LAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFA LCFVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGD LRNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFID QSFQESFAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMR DIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEEGEEP NYPLELLYCLYETQEDAFVRQALGRFPELALQRVRFCRMDVAVLSYCVRCCPAAQALR LISCRLVAAQEKKKKSLGKRLQASLGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHC KLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPD PQRGLQYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELS EQSLQELQAVKRAKPDLVITHPALDGHPQPPKELISTF		

Further analysis of the NOV18a protein yielded the following properties shown in Table 18B.

Table 18B. Protein Sequence Properties NOV18a	
Psort analysis:	0.7400 probability located in nucleus; 0.6000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 18C.

Table 18C. Geneseq Results for NOV18a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE04546	Human G-protein coupled receptor-2 (GCREC-2) protein - Homo sapiens, 891 aa. [WO200142288-A2, 14-JUN-2001]	1..676 210..891	671/682 (98%) 671/682 (98%)	0.0
AAU00023	Human activated T-lymphocyte associated sequence 2, ATLAS-2 - Homo sapiens, 1851 aa. [WO200114564-A2, 01-MAR-2001]	1..633 210..904	605/695 (87%) 610/695 (87%)	0.0
ABB11735	Human vasopressin receptor homologue, SEQ ID NO:2105 - Homo sapiens, 597 aa. [WO200157188-A2, 09-AUG-2001]	1..490 106..595	485/490 (98%) 485/490 (98%)	0.0
AAR33389	AII/AVPv2 receptor - Synthetic, 481 aa. [WO9305073-A, 18-MAR-1993]	193..670 1..480	322/481 (66%) 371/481 (76%)	e-174
AAM89960	Human immune/haematopoietic antigen SEQ ID NO:17553 - Homo sapiens, 329 aa. [WO200157182-A2, 09-AUG-2001]	1..274 9..282	265/274 (96%) 266/274 (96%)	e-151

5   In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18D.

154

Table 18D. Public BLASTP Results for NOV18a

Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC34689	SEQUENCE 3 FROM PATENT WO0114564 - Homo sapiens (Human), 1851 aa.	1..633 210..904	605/695 (87%) 610/695 (87%)	0.0
Q91WS2	HYPOTHETICAL 62.5 KDA PROTEIN - Mus musculus (Mouse), 556 aa (fragment).	107..659 1..554	390/557 (70%) 450/557 (80%)	0.0
Q63035	VASOPRESSIN RECEPTOR - Rattus norvegicus (Rat), 483 aa.	193..670 1..482	324/483 (67%) 372/483 (76%)	e-173
AAL12498	CRYOPYRIN - Homo sapiens (Human), 920 aa.	3..657 234..914	232/709 (32%) 355/709 (49%)	5e-94
AAL12497	CRYOPYRIN - Homo sapiens (Human), 1034 aa.	3..648 234..848	223/658 (33%) 344/658 (51%)	6e-93

PFam analysis predicts that the NOV18a protein contains the domains shown in the Table 18E.

Table 18E. Domain Analysis of NOV18a

Pfam Domain	NOV18a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 19.

The NOV19 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 19A.

Table 19A. NOV19 Sequence Analysis

	SEQ ID NO: 71	2686 bp
NOV19a, CG58626-01 DNA Sequence	CCGGCGGCGTCTCCACAGCATGAATTACCCGGGCCGCGGGTCCCCACGGAGCCCCGAG CATAACGGCCGAGGCGGCGGCGGCGGCGCCTGGGAGCTGGGCTCAGACGCGAGGCCAG CGTTCGGCGGCGGCGTCTGCTGCTTCGAGCACCTGCCCGGCGGGGACCCGGACGACGG CGACGTGCCCCTGGCCCTGCTGCGCGGGGAACCCGGGCTGCATTTGGCGCCGGGCACC GACGACCACAACCACCACCTCGCGCTGGACCCCTGCCTCAGTGACGAGAACTATGACT TCAGCTCCGCCGAGTCGGGCTCCTCGCTGCGCTACTACAGCGAGGGTGAGAGCGGCGG CGGCGGCAGCTCCTTGTCGCTGCACCCGCCGCAGCAGCCTCCGCTGGTCCCGACGAAC TCGGGGGGCGGCGGCGCGACAGGAGGGTCCCCCGGGGAAAGGAAACGTACCCGGCTTG GCGGCCCGGCGGCCCGGCACCGCTATGAGGTAGTGACGGAGCTGGGCCCGGAGGAGGT ACGCTGGTTCTACAAGGAGGACAAGAAGACCTGGAAGCCCTTCATCGGCTACGACTCG CTCCGCATCGAGCTCGCCTTCCGGACCCTGCTGCAGACCACGGGTGCCCGGCCCCAGG GCGGGGACCGGGACGGCGACCATGTGTGCTCCCCCACGGGCCCAGCCTCCAGTTCCGG AGAAGATGACGATGAGGACCGCGCCTGCGGCTTCTGCCAGAGTACGACGGGGCACGAG CCGGAGATGGTGGAGCTTGTGAACATCGAGCCTGTGTGCGTGCGGGGCGGCCTCTACG

155

	AGGTGGATGTGACCCAAGGAGAGTGCTACCCGGTGTACTGGAACCGTGCTGATAAAAT ACCAGTAATGCGTGGACAGTGGTTTATTGACGGCACTTGGCAGCCTCTAGAAGAGGAA GAAAGTAATTTAATTGAGCAAGAACATCTCAATTGTTTTAGGGGCCAGCAGATGCAGG AAAATTTCGATATTGAAGTGTCAAAATCCATAGATGGAAAAGATGCTGTTCATAGTTT CAAGTTGAGTCGAAACCATGTGGACTGGCACAGTGTGGATGAAGTATATCTTTATAGT GATGCAACAACATCTAAAATTGCAAGAACAGTTACCCAAAAACTGGGATTTTCTAAAG CATCAAGTAGTGGTACCAGACTTCATAGAGGTTATGTAGAAGAAGCCACATTAGAAGA CAAGCCATCACAGACTACCCATATTGTATTTGTTGTGCATGGCATTGGGCAGAAAATG GACCAAGGAAGAATTATCAAAAATACAGCTATGATGAGAGAAGCTGCAAGAAAAATAG AAGAAAGGCATTTTTCCAACCATGCAACACATGTTGAATTTCTGCCTGTTGAGTGGCG GTCAAAACTTACTCTTGATGGAGACACTGTTGATTCCATTACTCCTGACAAAGTACGA GGTTTAAGGGATATGCTGAACAGCAGTGCAATGGACATAATGTATTATACTAGTCCAC TTTATAGAGATGAACTAGTTAAAGGCCTTCAGCAAGAGCTGAATCGATTGTATTCCCT TTTCTGTTCTCGGAATCCAGACTTTGAAGAAAAAGGGGGTAAAGTCTCAATAGTATCA CATTCCTTGGGATGTGTAATTACTTATGACATAATGACTGGCTGGAATCCAGTTCGGC TGTATGAACAGTTGCTGCAAAAGGAAGAAGAGTTGCCTGATGAACGATGGATGAGCTA TGAAGAACGACATCTTCTTGATGAACTCTATATAACTAAACGACGGCTGAAGGAAATA GAAGAACGGCTTCACGGATTGAAAGCATCATCTATGACACAAACACCTGCCTTAAAAT TTAAGGTAGAGAATTTCTTCTGTATGGGATCCCCATTAGCAGTTTTCTTGGCGTTGCG TGGCATCCGCCCAGGAAATACTGGAAGTCAAGACCATATTTTGCCTAGAGAGATTTGT AACCGGTTACTAAATATTTTTCATCCTACAGATCCAGTGGCTTATAGATTAGAACCAT TAATACTGAAACACTACAGCAACATTTCACCTGTCCAGATCCACTGGTACAATACTTC AAATCCTTTACCTTATGAACATATGAAGCCAAGCTTTCTCAACCCAGCTAAAGAACCT ACCTCAGTTTCAGAGAATGAAGGCATTTCAACCATACCAAGCCCTGTGACCTCACCAG TTTTGTCCCGCCGACACTATGGAGAATCTATAACAAATATAGGCAAAGCAAGCATATT AGGTGCTGCTAGCATTGGAAAGGGACTTGGAGGAATGTTGTTCTCAAGATTTGGACGT TCATCTACAACACAGTCATCTGAAACATCAAAAGACTCAATGGAAGATGAGAAGAAGC CAGTTGCCTCACCTTCTGCTACCACCGTAGGGACACAGACCCTTCCACATAGCAGTTC TGGCTTCCTCGATTCTGCAGTGGAGTTGGATCACAGGATTGATTTTGAACTCAGAGAA GGCCTTGTGGAGAGCCGCTATTGGTCAGCTGTCACGTCGCATACTGCCTATTGGTCAT CCTTGGATGTTGCCCTTTTTCTTTTAACCTTCATGTATAAACATGAGCACGATGATGA TGCAAAACCCAATTTAGATCCAATCTG__ACTCTCTTGAAGGACATGAATGGCCTAAAA__ __CTGATTTTTTTTTTTCC__
	ORF Start: ATG at 20 **ORF Stop: TGA at 2636**
	SEQ ID NO: 72 **872 aa** **MW at 97063.4kD**
NOV19a, CG58626-01 Protein Sequence	MNYPGRGSPRSPEHNGRGGGGGAWELGSDARPAFGGGVCCFEHLPGGDPDDGDVPLAL LRGEPGLHLAPGTDDHNHHLALDPCLSDENYDFSSAESGSSLRYYSEGESGGGGSSLS LHPPQQPPLVPTNSGGGGATGGSPGERKRTRLGGPAARHRYEVVTELGPEEVRWFYKE DKKTWKPFIGYDSLRIELAFRTLLQTTGARPQGGDRDGDHVCSPTGPASSSGEDDDED RACGFCQSTTGHEPEMVELVNIEPVCVRGGLYEVDVTQGECYPVYWNRADKIPVMRGQ WFIDGTWQPLEEEESNLIEQEHLNCFRGQQMQENFDIEVSKSIDGKDAVHSFKLSRNH VDWHSVDEVYLYSDATTSKIARTVTQKLGFSKASSSGTRLHRGYVEEATLEDKPSQTT HIVFVVHGIGQKMDQGRIIKNTAMMREAARKIEERHFSNHATHVEFLPVEWRSKLTLD GDTVDSITPDKVRGLRDMLNSSAMDIMYYTSPLYRDELVKGLQQELNRLYSLFCSRNP DFEEKGGKVSIVSHSLGCVITYDIMTGWNPVRLYEQLLQKEEELPDERWMSYEERHLL DELYITKRRLKEIEERLHGLKASSMTQTPALKFKVENFFCMGSPLAVFLALRGIRPGN TGSQDHILPREICNRLLNIFHPTDPVAYRLEPLILKHYSNISPVQIHWYNTSNPLPYE HMKPSFLNPAKEPTSVSENEGISTIPSPVTSPVLSRRHYGESITNIGKASILGAASIG KGLGGMLFSRFGRSSTTQSSETSKDSMEDEKKPVASPSATTVGTQTLPHSSSGFLDSA VELDHRIDFELREGLVESRYWSAVTSHTAYWSSLDVALFLLTFMYKHEHDDDAKPNLD PI

Further analysis of the NOV19a protein yielded the following properties shown in Table 19B.

Table 19B. Protein Sequence Properties NOV19a	
PSort analysis:	0.4555 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV19a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 19C.

Table 19C. Geneseq Results for NOV19a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG64151	Arabidopsis thaliana gravitropism protein - Arabidopsis thaliana, 933 aa. [JP2001120279-A, 08-MAY-2001]	257..547 156..454	104/316 (32%) 156/316 (48%)	1e-38
AAM41595	Human polypeptide SEQ ID NO 6526 - Homo sapiens, 677 aa. [WO200153312-A1, 26-JUL-2001]	261..548 52..328	94/301 (31%) 138/301 (45%)	6e-25
AAB92643	Human protein sequence SEQ ID NO:10972 - Homo sapiens, 1000 aa. [EP1074617-A2, 07-FEB-2001]	119..608 226..664	132/524 (25%) 204/524 (38%)	2e-24
AAM39809	Human polypeptide SEQ ID NO 2954 - Homo sapiens, 615 aa. [WO200153312-A1, 26-JUL-2001]	274..548 3..266	90/288 (31%) 131/288 (45%)	4e-23
AAB93825	Human protein sequence SEQ ID NO:13636 - Homo sapiens, 694 aa. [EP1074617-A2, 07-FEB-2001]	404..608 227..449	76/229 (33%) 113/229 (49%)	6e-23

In a BLAST search of public sequence databases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19D.

157

Table 19D. Public BLASTP Results for NOV19a				
Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O46606	PHOSPHATIDIC ACID-PREFERRING PHOSPHOLIPASE A1 - Bos taurus (Bovine), 875 aa.	1..872 1..875	802/876 (91%) 829/876 (94%)	0.0
Q9C0F8	KIAA1705 PROTEIN - Homo sapiens (Human), 498 aa (fragment).	378..872 4..498	493/495 (99%) 494/495 (99%)	0.0
Q96LL2	CDNA FLJ25408 FIS, CLONE TST02965, HIGHLY SIMILAR TO BOS TAURUS PHOSPHATIDIC ACID-PREFERRING PHOSPHOLIPASE A1 MRNA - Homo sapiens (Human), 454 aa.	419..872 1..454	453/454 (99%) 454/454 (99%)	0.0
AAH18552	HYPOTHETICAL 27.3 KDA PROTEIN - Mus musculus (Mouse), 249 aa (fragment).	624..869 1..246	224/246 (91%) 236/246 (95%)	e-130
AAL32232	HYPOTHETICAL 85.1 KDA PROTEIN - Caenorhabditis elegans, 753 aa.	122..867 11..750	255/794 (32%) 374/794 (46%)	6e-91

PFam analysis predicts that the NOV19a protein contains the domains shown in the Table 19E.

Table 19E. Domain Analysis of NOV19a			
Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value
DUF203: domain 1 of 1	252..458	42/219 (19%) 105/219 (48%)	7.5
DDHD: domain 1 of 1	611..858	96/266 (36%) 236/266 (89%)	3.3e-116

Example 20.

The NOV20 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 20A.

<table>
<tr><td colspan="3">Table 20A. NOV20 Sequence Analysis</td></tr>
</table>

	SEQ ID NO: 73	773 bp
NOV20a, CG57597-01 DNA Sequence	GGTAAGGACACAAGATGCCAAATAGGGTAAGGAATGGTCCAGAAACCTGTGAACTCTG CATTGCAGGCATGCACCACCACTCCTGGCTAATTTTTTGTATTTTTAGTGCCATCGAA TCCGGCTCAAACCTTTTATTTCTCTTATGTAAAAGCTGTGTACTTCAGAAAAACATGT ACAGTTATCCCTGGCAGTGCCGGGGTGGGGTCTGCGCGGCCCTGGAGGCCTGGCCGGC CTTGCAGATCGCTGTGGAGAATGGCTTCGGGGGTGTGCACAGCCAGGAGAAGGCCAAG TGGCTGGGGGGTGCAGTGGAGGATTACTTCATGCGCAATGCTGACTTGGAGCTAGATG AGGTGGAAGACTTCCTTGGAGAGCTGTTGACCAACGAGTTTGATACAGTTGTGGAAGA CGGGAGTCTGCCCCAGGTGAGCCAGCAACTGCAGACCATGTTCCACCACTTCCAGAGG GGTGATGGGGCTGCTCTGAGGGAGATGGCCTCCTGCATCACTCAGAGAAAATGCAAGG TCACAGCCACTGCACTTAAGACAGCTAGAGAGACTGATGAGGATGAAGATGATGTGGA CAGTGTGGAAGAGATGGAGGTCACAGCTACGAATGATGGGGCTGCTACAGATGGGGTC TGCCCCCAGCCTGAACCCTCTGATCCAGACGCTCAGACTATTAAGGAAGAGGATATAG TGGAAGATGGCTGGACCATTGTCCGGAGAAAAAAATGAGTGGGGATGATTGGAAATGG CTTTGGGCCCTTATTTGCT	
	ORF Start: ATG at 15	ORF Stop: TGA at 732
	SEQ ID NO: 74	239 aa	MW at 26579.5kD
NOV20a, CG57597-01 Protein Sequence	MPNRVRNGPETCELCIAGMHHHSWLIFCIFSAIESGSNLLFLLCKSCVLQKNMYSYPW QCRGGVCAALEAWPALQIAVENGFGGVHSQEKAKWLGGAVEDYFMRNADLELDEVEDF LGELLTNEFDTVVEDGSLPQVSQQLQTMFHHFQRGDGAALREMASCITQRKCKVTATA LKTARETDEDEDDVDSVEEMEVTATNDGAATDGVCPQPEPSDPDAQTIKEEDIVEDGW TIVRRKK		

Further analysis of the NOV20a protein yielded the following properties shown in Table 20B.

<table>
<tr><td colspan="2">Table 20B. Protein Sequence Properties NOV20a</td></tr>
</table>

PSort analysis:	0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV20a protein against the Geneseq database, a proprietary

database that contains sequences published in patents and patent publication, yielded

5 several homologous proteins shown in Table 20C.

<table>
<tr><td colspan="5">Table 20C. Geneseq Results for NOV20a</td></tr>
</table>

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG81374	Human AFP protein sequence SEQ ID NO:266 - Homo sapiens, 191 aa. [WO200129221-A2, 26-APR-2001]	61..239 13..191	178/179 (99%) 178/179 (99%)	e-101
AAG57770	Arabidopsis thaliana protein	63..239 18..178	56/182 (30%) 94/182 (50%)	1e-13

	Arabidopsis thaliana, 184 aa. [EP1033405-A2, 06-SEP-2000]			
AAG57771	Arabidopsis thaliana protein fragment SEQ ID NO: 74487 - Arabidopsis thaliana, 156 aa. [EP1033405-A2, 06-SEP-2000]	74..239 1..150	52/171 (30%) 89/171 (51%)	2e-11

In a BLAST search of public sequence databases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20D.

Table 20D. Public BLASTP Results for NOV20a				
Protein Accession Number	Protein/Organism/Length	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q969E8	UNKNOWN (PROTEIN FOR MGC:20451) (PROTEIN FOR IMAGE:3953868) - Homo sapiens (Human), 191 aa.	61..239 13..191	178/179 (99%) 178/179 (99%)	e-101
Q9NAD8	Y51H4A.15 PROTEIN - Caenorhabditis elegans, 225 aa.	1..239 1..225	66/239 (27%) 122/239 (50%)	5e-23
Q06672	HIGHLY ACIDIC C-TERMINUS - Saccharomyces cerevisiae (Baker's yeast), 249 aa.	63..238 79..244	46/177 (25%) 82/177 (45%)	5e-11
Q9VBI0	CG14543 PROTEIN - Drosophila melanogaster (Fruit fly), 195 aa.	71..238 24..195	49/174 (28%) 81/174 (46%)	2e-10
Q9UUA9	HYPOTHETICAL HIGHLY ACIDIC C-TERMINUS PROTEIN - Schizosaccharomyces pombe (Fission yeast), 179 aa.	70..239 22..178	42/172 (24%) 83/172 (47%)	2e-06

PFam analysis predicts that the NOV20a protein contains the domains shown in the Table 20E.

5

Table 20E. Domain Analysis of NOV20a			
Pfam Domain	NOV20a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 21.

The NOV21 clone was analyzed, and the nucleotide and predicted polypeptide
sequences are shown in Table 21A.

Table 21A. NOV21 Sequence Analysis		
	SEQ ID NO: 75	7741 bp
NOV21a, CG57804-01 DNA Sequence	TTGTCTCTTTGTGTTTTCCAGACATTCTAAGTGAGACTGTCCACATCATCTAGGAAAA TGGTGGCCCTGTCCTTAAAGATTTGTGTGCGCCACTGCAACGTGGTGAAGACCATGCA GTTTGAACCATCTACAGCTGTGTACGATGCGTGTCGAGTCATTCGGGAACGGGTGCCT GAGGCACAAACTGGGCAAGCTTCTGACTATGGACTCTTTCTTTCGGATGAAGACCCGA GGAAAGGGATTTGGCTGGAAGCGGGCAGAACACTGGATTACTACATGTTGCGGAATGG GGATATTTTGGAATATAAAAAGAAACAGAGACCTCAGAAAATCCGGATGCTGGATGGA TCTGTGAAGACAGTGATGGTGGATGATTCCAAGACTGTGGGGGAGCTCCTGGTCACTA TTTGTAGCAGAATAGGAATAACAAATTATGAAGAATACTCCTTAATCCAAGAAACTAT TGAAGAAAGAAAGAGGAAGGAACGGGCACACTCAAAAAAGACAGGCACTGTTACGA GATGAGAGGAAAATGGAGAAGTTGAAGGCCAAGCTGCACACAGATGATGACCTAAATT GGCTGGATCACAGCCGAACATTCAGAGAACAAGGAGTAGATGAAAACGAAACGTTGCT GCTTAGACGGAAGTTCTTTTACTCTGATCAGAATGTAGATTCGAGAGACCCCGTGCAG CTGAACTTGCTTTATGTTCAGGCACGGGATGACATCCTGAATGGCTCTCACCCTGTCT CCTTCGAGAAAGCTTGTGAGTTTGGTGGATTTCAAGCCCAGATACAATTTGGACCTCA TGTGGAACATAAACACAAACCTGGATTTTTAGATCTGAAGGAATTCCTGCCCAAAGAA TATATCAAGCAGAGAGGAGCTGAAAAGAGGATCTTTCAGGAGCATAAGAACTGCGGAG AGATGAGTGAGATAGAAGCCAAGGTCAAGTACGTCAAACTCGCACGGTCCCTCCGCAC ATATGGCGTGTCCTTCTTCCTGGTGAAGGAGAAGATGAAAGGCAAGAACAAGCTGGTG CCTCGCCTGCTGGGGATCACCAAAGACTCGGTGATGCGCGTGGATGAGAAGACCAAGG AAGTGCTGCAGGAGTGGCCCCTCACCACCGTCAAGCGCTGGGCAGCCTCACCCAAGAG CTTCACACTGGATTTTGGGGAGTATCAGGAAAGCTACTATTCAGTACAAACCACCGAG GGAGAGCAGATATCCCAGCTGATTGCAGGCTACATTGACATCATCCTGAAAAAGGGAA CATACGTGACATCTGTGGGGTCTCCTCATTGCACTCCACATGGCTGGTGTTCTCTCAG TGACCAAACCACTTTTCCCGGCAGGTCCACCATCTTGCAGCAGCAGTTCAACCGGACC GGGAAGGCAGAGCACGGCTCAGTGGCGCTGCCGGCCGTGATGCGCTCGGGCTCCAGCG GGCCTGAGACCTTCAACGTTGGCAGCATGCCCTCGCCACAGCAGCAGGTCATGGTTGG GCAGATGCACCGAGGCCACATGCCGCCACTGACCTCAGCCCAGCAGGCCCTGATGGGG ACCATCAACACAAGCATGCACGCCGTCCAGCAGGCCCAGGATGATCTCAGTGAGCTCG ACTCGCTGCCACCTCTCGGCCAGGATATGGCATCTAGGGTATGGGTTCAGAACAAAGT CGACGAATCCAAACACGAAATCCATTCTCAAGTTGATGCTATCACGGCCGGAACGGCT TCAGTTGTTAACCTCACAGCTGGTGACCCTGCAGACACTGACTACACAGCTGTGGGAT GTGCGATCACCACTATTTCTTCCAACCTGACGGAGATGTCCAAGGGGTGTGAAGCTATT GGCCGCCCTCATGGATGATGAGGTGGGCAGCGGGGAGGACTTGCTCAGAGCTGCCAGG ACCCTCGCTGGGGCGGTGTCAGACTTGCTGAAAGCTGTGCAGCCTACTTCTGGAGAGC CTCGACAGACAGTTTTGACTGCTGCTGGCAGCATCGGACAAGCCAGTGGGGATCTTCT GAGACAGATTGGAGAGAATGAGACTGATGAGCGATTCCAGGATGTTTTAATGAGTTTG GCCAAAGCTGTTGCCAATGCAGCTGCCATGTTGGTACTAAAGGCAAAGAATGTTGCCC AAGTGGCCGAAGCACTGTCCTACAGAACAGGGTAATTGCTGCTGCCACCCAGTGTGC CCTCTCCACCTCCCAGCTTGTGGCATGTGCCAAGGTTGTGAGCCCCACTATTAGCTCC CCTGTGTGCCAGGAGCAGCTGATTGAAGCAGGGAAGCTGGTGGACCGCTCGGTGGAGA ACTGTGTCCGTGCCTGCCAGGCGGCCACTACCGATAGTGAGCTCCTGAAGCAGGTCAG CGCAGCGGCCAGCGTGGTCAGCCAGGCCCTCCATGATCTCCTGCAGCATGTGCGGCAG TTTGCCAGCCGAGGCGAGCCCATCGGCCGCTACGACCAGGCTACTGACACCATCATGT GTGTCACCGAGAGCATCTTCAGCTCCATGGGTGACGCTGGTGAAATGGTGCGCCAGGC GCGGGTTCTGGCCCAAGCCACATCAGACCTCGTCAATGCCATGAGGTCAGATGCAGAA GCCGAAATCGACATGGAGAATTCAAAGAAGCTCCTGGCAGCAGCAAAACTCTTAGCTG ACTCCACTGCTCGCATGGTGGAAGCTGCAAAGGGGGCTGCAGCCAACCCAGAGAATGA GGACCAGCAGCAAAGGCTGAGAGAAGCTGCAGAAGGCCTCCGGGTAGCAACCAACGCA GCTGCCCAGAATGCTATTAAGAAAAAAATTGTCAACCGACTGGAGGTTGCAGCCAAGC AGGCCGCAGCGGCAGCCACACAGACCATCGCCGCCTCCCAGAATGCAGCTGTTTCCAA CAAGAACCCTGCGGCCCAGCAGCAGCTGGTCCAGAGTTGCAAGGCAGTGGCTGATCAC ATCCCTCAGCTGGTCCAGGGAGTGAGGGGGAGCCAAGCTCAAGCTGAAGACCTGAGTG CCCAGCTGGCTCTCATCATCTCCAGCCAGAACTTCCTCCAGCCTGGAAGCAAGATGGT GTCCTCTGCCAAAGCCGCAGTGCCCACCGTGAGTGACCAGGCCGCAGCCATGCAGCTG AGCCAGTGTGCCAAGAACCTGGCCACCAGCTTGGCGGAGCTGCGTACCGCCTCGCAGA AGGCCCATGAAGCTTGTGGTCCGATGGAAATCGATTCAGCTCTGAATACGGTGCAGAC GCTTAAGAATGAACTGCAGGATGCCAAGATGGCAGCCGTGGAGAGCCAGCTGAAGCCA CTTCCAGGGGAAACGCTGGAAAAATGTGCTCAGGACCTGGGAAGCACATCCAAGGCGG TGGGCTCCTCCATGGCACAGCTGCTGACCTGTGCTGCTCAAGGCAACGAACACTACAC AGGGGTGGCTGCTAGAGAGACGGCCCAAGCTCTGAAAACACTGGCCCAGGCCGCCCGT GGAGTGGCTGCATCGACAACCGACCCCGCGGCCGCCCATGCCATGTTAGATTCTGCTC GAGACGTGATGGAGGGCTCCGCCATGCTCATTCAAGAGGCCAAGCAGGCCCTGATTGC ACCTGGAGATGCAGAGCGTCAACAAAGACTGGCTCAGGTGGCTAAAGCCGTCTCACAC TCCTTGAATAACTGCGTAAATTGCCTCCCTGGGCAGAAGGATGTGGACGTGGCCTTGA

161

	AGAGCATCGGGGAGTCCAGCAAGAAGCTGCTTGTGGATTCGCTACCTCCAAGCACGAA GCCTTTCCAGGAAGCCCAGAGTGAACTGAACCAGGCAGCAGCTGATCTGAACCAGTCT GCTGGGGAAGTGGTCCATGCCACCCGGGGCCAGAGTGGAGAGTTGGCTGCAGCCTCTG GAAAGTTCAGTGATGATTTTGGTGAATTCCTCGATGCTGGCATTGAGATGGCTGGCCA AGCTCAGACAAAAGAAGACCAGATCCAAGTGATAGGGAACCTCAAGAATATCTCGATG GCATCCAGCAAGCTGCTGTTAGCTGCCAAGTCTCTCTCTGTAGATCCAGGAGCTCCCA ATGCGAAAAATCTCCTGGCTGCAGCTGCAAGAGCTGTGACAGAGAGCATCAATCAACT CATCACTCTGTGTACCCAACAAGCTCCGGGCCAGAAAGAGTGCGATAATGCCCTGCGG GAGCTCGAGACTGTGAAGGGGATGTTGGACAATCCTAATGAACCTGTTAGTGACCTCT CTTACTTTGACTGCATTGAGAGTGTGATGGAAAACTCCAAGGTTCTGGGTGAATCGAT GGCAGGGATTTCACAGAATGCCAAGACCGGAGACCTCCCTGCCTTTGGGGAATGTGTG GGGATTGCATCCAAGGCTCTCTGTGGGCTGACAGAGGCTGCAGCCCAGGCTGCATACT TGGTTGGCATCTCTGATCCAAACAGCCAGGCAGGCCACCAGGGCCTGGTGGACCCCAT CCAGTTTGCCAGGGCTAACCAGGCCATCCAGATGGCATGCCAGAACTTGGTGGACCCT GGCAGCAGCCCATCACAGGTCCTGTCAGCCGCCACAATTGTTGCCAAGCACACGTCAG CCTTGTGCAATGCCTGCCGCATCGCCTCATCCAAGACGGCCAACCCAGTAGCCAAGAG GCACTTCGTCCAGTCAGCCAAGGAAGTCGCCAACAGCACTGCCAACCTGGTGAAGACC ATCAAGGCCCTGGATGGGGATTTCTCTGAAGACAACCGCAATAAGTGTCGCATCGCCA CCGCACCCTTGATTGAAGCTGTGGAGAACCTGACAGCGTTCGCCTCAAACCCTGAGTT TGTCAGCATTCCTGCCCAGATCAGCTCCGAGGGTTCCCAGGCACAGGAACCAATCCTG GTCTCAGCCAAGACCATGCTGGAGAGTTCATCGTACCTCATTCGCACTGCACGCTCTC TGGCCATCAACCCCAAAGACCCACCCACCTGGTCTGTACTGGCTGGACATTCCCATAC AGTGTCCGACTCCATCAAGAGTCTCATCACTTCTATCAGGGACAAGGCCCCTGGACAG AGGGAGTGTGATTACTCCATCGATGGCATCAACCGGTGCATCCGGGACATCGAGCAGG CCTCGCTGGCCGCCGTCAGCCAGAGCCTGGCCACGAGGGACGACATCTCTGTGGAGGC CCTGCAGGAGCAGCTGACTTCGGTGGTCCAGGAAATCGGACACCTTATCGATCCCATC GCCACAGCGGCTCGGGGAGAAGCAGCTCAGCTGGGACATAAGGTGACACAACTGGCAA GCTATTTTGAGCCCTTGATCTTAGCCGCAGTTGGTGTGGCCTCCAAGATTCTTGATCA TCAGCAGCAGATGACGGTGCTGGACCAGACCAAGACTCTCGCAGAGTCTGCCTTGCAG ATGTTGTATGCAGCCAAAGAAGGTGGCGGAAACCCCAAGGCACAACACACCCATGACG CCATCACAGAGGCCGCCCAGTTGATGAAGGAAGCCGTGGATGACATCATGGTGACGCT GAACGAAGCTGCCAGTGAAGTGGGGCTGGTTGGGGGCATGGTGGACGCCATTGCAGAA GCCATGAGCAAGCTGGATGAAGGCACTCCTCCAGAACCAAAGGGAACATTTGTCGACT ATCAGACGACTGTGGTTAAATACTCCAAAGCCATTGCGGTGACAGCTCAGGAAATGAT GACTAAGTCGGTTACTAACCCGGAGGAGTTGGGGAGGACTGGCTTCACAAATGACCAGT GACTATGGGCACCTGGCTTTCCAGGGCCAGATGGCAGCAGCCACGGCGGAACCAGAGG AGATCGGATTCCAGATTCGCACTCGTGTGCAGGACCTGGGCCACGGCTGTATCTTCCT GGTGCAGAAGGCAGGGGCCCTCCAGGTCTGCCCCACAGACAGCTACACCAAGAGGGAG CTGATCGAATGCGCCCGTGCCGTCACGGAAAAGGTCTCCTTGGTGCTCTCGGCTCTCC AGGCCGGGAACAAAGGAACCCAGGCATGCATTACAGCCGCCACCGCTGTGTCTGGGAT CATTGCCGACCTGGACACCACCATTATGTTTGCAACAGCGGGGCAGCTGAATGCAGAG AACAGTGAGACCTTCGCAGACCACAGGGAGAACATTCTCAAGACGGCCAAGGCCTTGG TAGAAGACACGAAACTACTTGTGTCAGGAGCTGCGTCCACTCCTGACAAGCTGGCCCA GGCGGCCCAGTCCTCAGCAGCCACCATCACCCAGCTCGCAGAAGTGGTCAAGCTGGGG GCAGCCAGCCTGGGCTCCGACGACCCCGAGACCCAGGTGGATTTGATCAATGCCATCA AAGATGTGGCCAAGGCCCTTTCTGATCTCATCAGTGCTACCAAGGGAGCTGCCAGCAA GCCAGTGGACGACCCTTCCATGTACCAGCTCAAGGGGGCTGCCAAGGTGATGGTGACC AATGTCACCTCGCTCCTCAAGACTGTAAAGGCAGTGGAGGATGAGGCCACCCGGGGCA CCAGGGCGCTTGAGGCCACAATTGAATGCATAAAGCAGGAGCTTACGGTGTTCCAGTC AAAAGACGTACCTGAAAAGACATCATCACCTGAAGAATCCATAAGGATGACGAAAGGC ATCACCATGGCAACAGCCAAAGCCGTGGCAGCTGGGAACTCATGTAGACAGGAGGACG TGATTGCTACTGCCAACCTGAGCCGGAAAGCCGTGTCAGATATGTTGACGGCTTGCAA GCAAGCATCCTTCCACCCCGATGTCAGTGACGAGGTGAGAACCAGAGCCTTGCGTTTC GGGACGGAGTGCACCCTTGGCTACTTGGACCTCCTGGAGCACGTCTTGGTGATTCTTC AGAAACCAACCCCAGAATTCAAGCAGCAGCTGGCCGCTTTCTCCAAGCGAGTCGCCGG CGCTGTGACAGAGCTCATCCAGGCGGCGGAAGCCATGAAAGGAACAGAGTGGGTGGAT CCAGAAGACCCAACTGTCATTGCAGAAACAGAGTTACTGGGGGCTGCAGCATCCATCG AAGCTGCTGCTAAGAAGTTAGAGCAACTGAAGCCAAGAGCAAAACCAAAACAAGCGGA TGAGACCCTGGACTTTGAGGAACAGATCTTGGAAGCTGCTAAATCCATTGCTGCTGCC ACAAGCGCCCTGGTCAAATCGGCCTCAGCAGCCCAGAGGGAGCTGGTGGCCCAAGGAA AGGTGGGCTCCATCCCTGCCAATGCTGCAGACGACGGACAGTGGTCACAGGGGCTGAT TTCTGCTGCCCGGATGGTGGCGGCTGCGACCAGCAGTCTCTGTGAGGCGGCCAATGCC TCCGTTCAGGGACACGCCAGCGAGGAGAAGCTCATCTCATCTGCCAAGCAGGTCGCCG CTTCCACGGCTCAGCTGCTGGTGGCCTGCAAGGTGAAGGCCGACCAGGATTCAGAGGC CATGAGGCGGCTACAGGCGGCAGGAAATGCTGTGAAAAGGCCTCAGACAATCTTGTC CGTGCAGCCCAGAAGGCAGCTTTTGGCAAAGCTGATGACGACGATGTTGTAGTGGAAA CCAAGTTTGTGGGGGGCATTGCTCAGATCATCGCCGCCCAGGAAGAAATGCTAAAGAA AGAGCGAGAACTGGAAGAAGCAAGGAAAAAACTGGCCCAAATCCGCCAGCAGCAGTAT AAGTTTTTACCCACCGAGCTGAGGGAAGATGAGGGCTAAAGGTGCGAGCCCAGATGGC GAGCCCCAGGGGATGGCCCTGGCTGAA
	ORF Start: ATG at 58	**ORF Stop: TAA at 7693**
	SEQ ID NO: 76	2545 aa	MW at 271692.8kD

| NOV21a, CG57804-01 Protein Sequence | MVALSLKICVRHCNVVKTMQFEPSTAVYDACRVIRERVPEAQTGQASDYGLFLSDEDP RKGIWLEAGRTLDYYMLRNGDILEYKKKQRPQKIRMLDGSVKTVMVDDSKTVGELLVT ICSRIGITNYEEYSLIQETIEEKKEEGTGTLKKDRTLLRDERKMEKLKAKLHTDDDLN WLDHSRTFREQGVDENETLLLRRKFFYSDQNVDSRDPVQLNLLYVQARDDILNGSHPV SFEKACEFGGFQAQIQFGPHVEHKHKPGFLDLKEFLPKEYIKQRGAEKRIFQEHKNCG EMSEIEAKVKYVKLARSLRTYGVSFFLVKEKMKGKNKLVPRLLGITKDSVMRVDEKTK EVLQEWPLTTVKRWAASPKSFTLDFGEYQESYYSVQTTEGEQISQLIAGYIDIILKKG TYVTSVGSPHCTPHGWCSLSDQTTFPGRSTILQQQFNRTGKAEHGSVALPAVMRSGSS GPETFNVGSMPSPQQQVMVGQMHRGHMPPLTSAQQALMGTINTSMHAVQQAQDDLSEL DSLPPLGQDMASRVWVQNKVDESKHEIHSQVDAITAGTASVVNLTAGDPADTDYTAVG CAITTISSNLTEMSKGVKLLAALMDDEVGSGEDLLRAARTLAGAVSDLLKAVQPTSGE PRQTVLTAAGSIGQASGDLLRQIGENETDERFQDVLMSLAKAVANAAAMLVLKAKNVA QVAEDTVLQNRVIAAATQCALSTSQLVACAKVVSPTISSPVCQEQLIEAGKLVDRSVE NCVRACQAATTDSELLKQVSAAASVVSQALHDLLQHVRQFASRGEPIGRYDQATDTIM CVTESIFSSMGDAGEMVRQARVLAQATSDLVNAMRSDAEAEIDMENSKKLLAAAKLLA DSTARMVEAAKGAAANPENEDQQQRLREAAEGLRVATNAAAQNAIKKKIVNRLEVAAK QAAAAATQTIAASQNAAVSNKNPAAQQQLVQSCKAVADHIPQLVQGVRGSQAQAEDLS AQLALIISSQNFLQPGSKMVSSAKAAVPTVSDQAAAMQLSQCAKNLATSLAELRTASQ KAHEACGPMEIDSALNTVQTLKNELQDAKMAAVESQLKPLPGETLEKCAQDLGSTSKA VGSSMAQLLTCAAQGNEHYTGVAARETAQALKTLAQAARGVAASTTDPAAAHAMLDSA RDVMEGSAMLIQEAKQALIAPGDAERQQRLAQVAKAVSHSLNNCVNCLPGQKDVDVAL KSIGESSKKLLVDSLPPSTKPFQEAQSELNQAAADLNQSAGEVVHATRGQSGELAAAS GKFSDDFGEFLDAGIEMAGQAQTKEDQIQVIGNLKNISMASSKLLLAAKSLSVDPGAP NAKNLLAAAARAVTESINQLITLCTQQAPGQKECDNALRELETVKGMLDNPNEPVSDL SYFDCIESVMENSKVLGESMAGISQNAKTGDLPAFGECVGIASKALCGLTEAAAQAAY LVGISDPNSQAGHQGLVDPIQFARANQAIQMACQNLVDPGSSPSQVLSAATIVAKHTS ALCNACRIASSKTANPVAKRHFVQSAKEVANSTANLVKTIKALDGDFSEDNRNKCRIA TAPLIEAVENLTAFASNPEFVSIPAQISSEGSQAQEPILVSAKTMLESSSYLIRTARS LAINPKDPPTWSVLAGHSHTVSDSIKSLITSIRDKAPGQRECDYSIDGINRCIRDIEQ ASLAAVSQSLATRDDISVEALQEQLTSVVQEIGHLIDPIATAARGEAAQLGHKVTQLA SYFEPLILAAVGVASKILDHQQQMTVLDQTKTLAESALQMLYAAKEGGGNPKAQHTHD AITEAAQLMKEAVDDIMVTLNEAASEVGLVGGMVDAIAEAMSKLDEGTPPEPKGTFVD YQTTVVKYSKAIAVTAQEMMTKSVTNPEELGGLASQMTSDYGHLAFQGQMAAATAEPE EIGFQIRTRVQDLGHGCIFLVQKAGALQVCPTDSYTKRELIECARAVTEKVSLVLSAL QAGNKGTQACITAATAVSGIIADLDTTIMFATAGTLNAENSETFADHRENILKTAKAL VEDTKLLVSGAASTPDKLAQAAQSSAATITQLAEVVKLGAASLGSDDPETQVDLINAI KDVAKALSDLISATKGAASKPVDDPSMYQLKGAAKVMVTNVTSLLKTVKAVEDEATRG TRALEATIECIKQELTVFQSKDVPEKTSSPEESIRMTKGITMATAKAVAAGNSCRQED VIATANLSRKAVSDMLTACKQASFHPDVSDEVRTRALRFGTECTLGYLDLLEHVLVIL QKPTPEFKQQLAAFSKRVAGAVTELIQAAEAMKGTEWVDPEDPTVIAETELLGAAASI EAAAKKLEQLKPRAKPKQADETLDFEEQILEAAKSIAAATSALVKSASAAQRELVAQG KVGSIPANAADDGQWSQGLISAARMVAAATSSLCEAANASVQGHASEEKLISSAKQVA ASTAQLLVACKVKADQDSEAMRRLQAAGNAVKRASDNLVRAAQKAAFGKADDDDVVVE TKFVGGIAQIIAAQEEMLKKERELEEARKKLAQIRQQQYKFLPTELREDEG |

Further analysis of the NOV21a protein yielded the following properties shown in Table 21B.

Table 21B. Protein Sequence Properties NOV21a	
PSort analysis:	0.5964 probability located in mitochondrial matrix space; 0.3037 probability located in mitochondrial inner membrane; 0.3037 probability located in mitochondrial intermembrane space; 0.3037 probability located in mitochondrial outer membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV21a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 21C.

Table 21C. Geneseq Results for NOV21a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB41087	Human ORFX ORF851 polypeptide sequence SEQ ID NO:1702 - Homo sapiens, 2541 aa. [WO200058473-A2, 05-OCT-2000]	1..2543 1..2540	1913/2546 (75%) 2231/2546 (87%)	0.0
AAM39312	Human polypeptide SEQ ID NO 2457 - Homo sapiens, 1165 aa. [WO200153312-A1, 26-JUL-2001]	1381..2545 1..1165	1161/1165 (99%) 1163/1165 (99%)	0.0
AAM79794	Human protein SEQ ID NO 3440 - Homo sapiens, 1177 aa. [WO200157190-A2, 09-AUG-2001]	1378..2545 10..1177	1156/1168 (98%) 1160/1168 (98%)	0.0
AAM41098	Human polypeptide SEQ ID NO 6029 - Homo sapiens, 1177 aa. [WO200153312-A1, 26-JUL-2001]	1378..2545 10..1177	1156/1168 (98%) 1160/1168 (98%)	0.0
AAM41079	Human polypeptide SEQ ID NO 6010 - Homo sapiens, 1177 aa. [WO200153312-A1, 26-JUL-2001]	1378..2545 10..1177	1156/1168 (98%) 1160/1168 (98%)	0.0

In a BLAST search of public sequence databases, the NOV21a protein was found to have homology to the proteins shown in the BLASTP data in Table 21D.

Table 21D. Public BLASTP Results for NOV21a				
Protein Accession Number	Protein/Organism/Length	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Y490	Talin - Homo sapiens (Human), 2541 aa.	1..2543 1..2540	1910/2546 (75%) 2230/2546 (87%)	0.0
P26039	Talin – Mus musculus (Mouse), 2541 aa.	1..2543 1..2540	1907/2546 (74%) 2230/2546 (86%)	0.0
Q9UPX3	KIAA1027 PROTEIN - Homo	853..2543 1..1694	1262/1694 (74%) 1483/1694 (87%)	0.0

164

	(fragment).			
Q9VSL8	CG6831 PROTEIN (TALIN) - Drosophila melanogaster (Fruit fly), 2836 aa.	1..2532 1..2534	1197/2563 (46%) 1707/2563 (65%)	0.0
Q9Y4G6	KIAA0320 PROTEIN - Homo sapiens (Human), 949 aa (fragment).	1597..2545 1..949	947/949 (99%) 948/949 (99%)	0.0

PFam analysis predicts that the NOV21a protein contains the domains shown in the Table 21E.

Table 21E. Domain Analysis of NOV21a			
Pfam Domain	NOV21a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ubiquitin: domain 1 of 1	64..88	8/27 (30%) 20/27 (74%)	4.3
Band_41: domain 1 of 1	123..316	67/211 (32%) 172/211 (82%)	1.3e-92
IRS: domain 1 of 1	312..404	19/109 (17%) 46/109 (42%)	1.2
I_LWEQ: domain 1 of 5	674..768	31/98 (32%) 59/98 (60%)	11
transport_prot: domain 1 of 1	667..814	24/182 (13%) 88/182 (48%)	10
I_LWEQ: domain 2 of 5	852..894	18/47 (38%) 31/47 (66%)	2.4e+02
Vinculin: domain 1 of 1	860..903	12/48 (25%) 30/48 (62%)	1.3
I_LWEQ: domain 3 of 5	925..984	21/62 (34%) 37/62 (60%)	5.9e+04
TP_methylase: domain 1 of 1	861..1036	26/226 (12%) 105/226 (46%)	8
Apolipoprotein: domain 1 of 1	981..1229	48/288 (17%) 141/288 (49%)	3.5
CAP: domain 1 of 1	917..1354	94/557 (17%) 209/557 (38%)	4.4

I_LWEQ: domain 4 of 5	1529..1545	10/17 (59%) 13/17 (76%)	56
STAT: domain 1 of 1	1660..1821	35/211 (17%) 95/211 (45%)	8.2
LEA: domain 1 of 1	1768..1834	15/76 (20%) 42/76 (55%)	7
Histone_HNS: domain 1 of 1	2232..2356	29/143 (20%) 63/143 (44%)	3.7
I_LWEQ: domain 5 of 5	2345..2536	100/202 (50%) 183/202 (91%)	2e-101

Example 22.

The NOV22 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 22A.

Table 22A. NOV22 Sequence Analysis		
	SEQ ID NO: 77	2214 bp
NOV22a, CG57551-01 DNA Sequence	ATTCCTCCCTGCCCCTCGTGCAGCCGCTGCC**ATG**GCCCAGACACTGCAGATGGAGATC CCGAACTTCGGCAACAGCATCCTGGAGTGCCTCAATGAACAGCGGCTGCAGGGCCTGT ACTGTGACGTGTCAGTGGTGGTCAAGGGCCATGCCTTCAAGGCCCACCGGGCCGTGCT TGCTGCCAGCAGCTCCTACTTCCGGGACCTGTTCAACAACAGCCGCAGCGCCGTGGTG GAGCTGCCGGCGGCTGTGCAGCCCCAGTCTTTCCAGCAGATCCTCAGCTTCTGCTACA CGGGGCCGGCTGAGCATGAACGTGGGCGACCAGTTCCTGCTCATGTACACGGCTGGCTT CCTGCAGATCCAGGAGATCATGGAGAAGGGCACCGAGTTCTTCCTCAAGGTGAGCTCC CCGAGCTGCGACTCCCAGGGCCTGCATGCGGAGGAGGCCCCATCGTCGGAGCCCCAGA GCCCCGTGGCGCAGACATCGGGCTGGCCAGCCTGTAGCACCCCGCTGCCCCTCGTGTC GCGGGTGAAGACGGAGCAGCAGGAGTCGGACTCCGTGCAGTGCATGCCCGTGGCCAAG CGGCTGTGGGACAGTGGCCAGAAGGAGGCTGGGGGCGGCGGCAATGGCAGCCGCAAGA TGGCCAAGTTCTCCACGCCGGACCTGGCTGCCAACCGGCCTCACCAGCCCCGCCACC CCAACAGGCTCCGGTGGTGGCAGCAGCCCAGCCCGCCGTGGCTGCGGGAGCAGGGCAG CCAGCCGGTGGGGTGGCAGCAGCAGGGGGTGTGGTGAGTGGGCCCAGCACGTCGGAGC GGACCAGCCCAGGCACCTCAAGCGCCTACACCAGCGACAGCCCTGGCTCCTACCACAA TGAGGAGGACGAGGAGGAGGATGGTGGCGAGGAGGGCATGGATGAGCAGTACCGGCAG ATCTGCAACATGTACACCATGTACAGCATGATGAACGTCGGCCAGACAGCCGAGAAGG TGGAGGCCCTCCCGGAGCAGGTAGCCCCCGAGTCCCGAAATCGCATCCGGGTTCGGCA AGACCTGGCGTCTCTCCCGGCTGAACTTATCAACCAGATTGGGAACCGCTGCCACCCC AAGCTCTACGACGAGGCGCACCCCTCTGAGAAGCTGGACGCTGGTGACAGGCACCAACG TGTACATCACAAGGGCGCAGCTGATGAACTGCCACGTCAGCGCAGGCACGCGGCACAA GGTCCTACTGCGGCGGCTCCTGGCCTCCTTCTTTGACCGGAACACGCTGGCCAACAGC TGCGGCACCGGCATCCGCTCTTCTACCAACGATCCCCGTCGGAAGCCCCTGGACAGCC GCGTGCTCCACGCTGTCAAGTACTACTGCCAGAACTTCGCCCCCAACTTCAAGGAGAG CGAGATGAATGCCATCGCGGCCGACATGTGCACCAACGCCCGCCGCGTCGTGCGCAAG AGCTGGATGCCCAAGGTCAAGGTGCTCAAGGCTGAGGATGACGCCTACACCACCTTCA TCAGTGAAACGGGCAAGATCGAGCCGGACATGATGGGTGTGGAGCATGGCTTCGAGAC CGCCAGCCACGAGGGCGAGGCGGGCGGGTCCCATCGCTGAAGCCCTGCAG**TAA**CCCGCCCAG CCTCCCGCGGGGCCGCACACTTCCCCTCCCAACACACACACACACCTGCCATCTTGGT CATGAGCTACTGTCTGTCCCTCCCCAGGACCCGCGGTGGGTGCTGCATGTTCCCGGCC CTCTGCCCCTCCTGTCCTACCCCCTTTCCCCACCGAGAGCTGGGCCGGGAGAGGACCG CAGGGCAGGTGGCGTGAGGTCCGTGTTGCCTTCTTTAACACACACTCGTGCAGTGGGG GAGTTCTGGCTCCCCAACCTAACCCCTAGCCGTCATCTCCACACTCACCAGGCCCACC AGGGGAGGGGGCTGGCCTGGGGGTCTTGGGAAGGCCCCTCCCCAGGCCTTAGGCCACC TCGCGGAAGCCTTCAGCCTCCGCCCCTCACTGCAGCCCCTTGGGACTTGAGGGGGGCC CCAGGGGTTCTCAGGACCCCTCCCACCACCTCCCAGTGCTTCCACGTCTCCAAAAGCG CCTTCCTGTCACCCTCGTCTATCCCTGCGCCTGGGGGCTGGGGTAGGCAGGCCGTGG GGACTACCCATTTTATAGCTGGGGAAACAGGCTCCGAGAAATTGCACAACCGACCTCA GGTGGCCGGC
	ORF Start: ATG at 32	ORF Stop: TAA at 1613
	SEQ ID NO: 78	527 aa	MW at 57283.8kD

166

| NOV22a,
CG57551-01 Protein Sequence | MAQTLQMEIPNFGNSILECLNEQRLQGLYCDVSVVVKGHAFKAHRAVLAASSSYFRDL
FNNSRSAVVELPAAVQPQSFQQILSFCYTGRLSMNVGDQFLLMYTAGFLQIQEIMEKG
TEFFLKVSSPSCDSQGLHAEEAPSSEPQSPVAQTSGWPACSTPLPLVSRVKTEQQESD
SVQCMPVAKRLWDSGQKEAGGGGNGSRKMAKFSTPDLAANRPHQPPPPQQAPVVAAAQ
PAVAAGAGQPAGGVAAAGGVVSGPSTSERTSPGTSSAYTSDSPGSYHNEEDEEEDGGE
EGMDEQYRQICNMYTMYSMMNVGQTAEKVEALPEQVAPESRNRIRVRQDLASLPAELI
NQIGNRCHPKLYDEGDPSEKLELVTGTNVYITRAQLMNCHVSAGTRHKVLLRRLLASF
FDRNTLANSCGTGIRSSTNDPRRKPLDSRVLHAVKYYCQNFAPNFKESEMNAIAADMC
TNARRVVRKSWMPKVKVLKAEDDAYTTFISETGKIEPDMMGVEHGFETASHEGEAGPI
AEALQ |

Further analysis of the NOV22a protein yielded the following properties shown in Table 22B.

Table 22B. Protein Sequence Properties NOV22a	
PSort analysis:	0.6000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted .

A search of the NOV22a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 22C.

Table 22C. Geneseq Results for NOV22a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB41621	Human ORFX ORF1385 polypeptide sequence SEQ ID NO:2770 - Homo sapiens, 228 aa. [WO200058473-A2, 05-OCT-2000]	300..527 1..228	228/228 (100%) 228/228 (100%)	e-131
ABB17117	Human nervous system related polypeptide SEQ ID NO 5774 - Homo sapiens, 190 aa. [WO200159063-A2, 16-AUG-2001]	409..501 1..93	64/94 (68%) 73/94 (77%)	7e-29
AAG78615	Human zinc finger transcription factor BioZFTF45 - Homo sapiens, 413 aa. [CN1299825-A, 20-JUN-2001]	5..159 7..170	62/164 (37%) 92/164 (55%)	2e-25
AAY73351	HTRM clone 1484257 protein	7..291 1..277	83/291 (28%) 124/291 (42%)	8e-18

	[WO9957144-A2, 11-NOV-1999]			
AAM41058	Human polypeptide SEQ ID NO 5989 - Homo sapiens, 804 aa. [WO200153312-A1, 26-JUL-2001]	7..291 2..271	84/295 (28%) 123/295 (41%)	2e-17

In a BLAST search of public sequence databases, the NOV22a protein was found to have homology to the proteins shown in the BLASTP data in Table 22D.

Table 22D. Public BLASTP Results for NOV22a				
Protein Accession Number	Protein/Organism/Length	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96RE7	NAC1 PROTEIN - Homo sapiens (Human), 527 aa.	1..527 1..527	526/527 (99%) 526/527 (99%)	0.0
O35260	NAC-1 PROTEIN - Rattus norvegicus (Rat), 514 aa.	1..527 1..514	462/530 (87%) 475/530 (89%)	0.0
Q9CZ72	4930511N13RIK PROTEIN - Mus musculus (Mouse), 514 aa.	1..527 1..514	462/530 (87%) 476/530 (89%)	0.0
Q96BF6	SIMILAR TO RIKEN CDNA 0610020I02 GENE - Homo sapiens (Human), 587 aa.	1..501 1..478	289/522 (55%) 335/522 (63%)	e-140
AAH22103	RIKEN CDNA 0610020I02 GENE - Mus musculus (Mouse), 586 aa.	1..485 1..459	281/502 (55%) 327/502 (64%)	e-139

PFam analysis predicts that the NOV22a protein contains the domains shown in the Table 22E.

Table 22E. Domain Analysis of NOV22a			
Pfam Domain	NOV22a Match Region	Identities/ Similarities for the Matched Region	Expect Value
BTB: domain 1 of 1	14..124	40/143 (28%) 88/143 (62%)	6.2e-23

5   Example 23.

The NOV23 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 23A.

Table 23A. NOV23 Sequence Analysis		
	SEQ ID NO: 79	1497 bp
NOV23a, CG57411-01 DNA Sequence	ATGGCCACTGCACAGGTGGAACTGGTGCAGGGTGGTCCCCGGGCTCCAGTAGGGGAGA AGCTGGAGCTCGTCCTGTCGAACCTGCAGGCAGACGTCCTGGAGTTGCTGCTGGAGTT TGTCTACACGGGCTCCCTGGTCATCGACTCGGCCAACGCCAAGACACTGCTGGAGGCG GCCAGCAAGTTCCAGTTCCACACCTTCTGCAAAGTCTGCGTGTCCTTTCTCGAGAAGC AGCTGACGGCCAGCAACTGCCTGGGCGTGCTGGCCATGGCCGAGGCCATGCAGTGCAG CGAGCTCTACCACATGGCCAAGGCCTTCGCGCTGCAGATCTTCCCCGAGGTGGCCGCC CAGGAGGAGATCCTCAGCATCTCCAAGGACGACTTCATCGCCTACGTCTCCAACGACA GCCTCAACACCAAGGCTGAGGAGCTGGTGTACGAGACAGTCATCAAGTGGATCAAGAA GGACCCCGCGACACGCACACAGCTGCAGTACGCGGCTGAGCTCCTGGCCGTGGTCCGC CTCCCCTTCATCCACCCCAGCTACCTGCTCAATGTGGTTGACAATGAAGAGCTGATCA AGTCATCAGAAGCCTGCCGGGACCTGGTGAACGAGGCCAAACGCTACCATATGCTGCC CCACGCCCGCCAGGAGATGCAGACGCCCCGAACCCGGCCGCGCGTCCCTGCAGGTGTG GCTGAGGTCATCGTCTTGGTTGGGGGCCGTCAGATGGTGGGGATGACCCAGCGCTCGC TGGTGGCCGTCACCTGCTGGAACCCGCAGAACAACAAGTGGTACCCCTTGGCCTCGCT GCCCTTCTATGACCGCGAGTTCTTCAGTGTAGTGAGTGCAGGGGACAACATCTACCTC TCAGGTGGGATGGAATCAGGGGTGACGCTGGCTGATGTCTGGTGCTACATGTCCCTGC TTGATAACTGGAACCTCGTCTCCAGAATGACAGTCCCCCGCTGTCGGCACAATAGCCT CGTCTACGATGGGAAGATTTACACCCTCGGGGGACTTGGCGTGGCAGGCAACGTGGAC CACGTGGAGGTCCCTGCAGGTGTGGCTGAGGTCATCGTCTTGGTTGGGGGCCGTCAGA TGGTGGGGATGACCCAGCGCTCGCTGGTGGCCGTCACCTGCTGGAACCCGCAGAACAA CAAGTGGTACCCCTTGGCCTCGCTGGGTGGGATGGAATCAGGGGTGACGCTGGCTGAT GTCTGGTGCTACATGTCCCTGCTTGATAACTGGAACCTCGTCTCCAGAATGACAGTCC CCCGCTGTCGGCACAATAGCCTCGTCTACGATGGGAAGATTTACACCCTCGGGGGACT TGGCGTGGCAGGCAACGTGGACCACGTGGAGGCCTACGAGCCCACAACCAACACATGG ACCCTCCTCCCCCACATGCCCTGCCCTGTGTTCAGACACGGCTGCGTCGTGATAAAGA AATATATTCAAAGCGGCTGACATCAGCAGAAAGCCCACGATAAGACT
	ORF Start: ATG at 1	ORF Stop: TGA at 1468
	SEQ ID NO: 80	489 aa	MW at 54208.2kD
NOV23a, CG57411-01 Protein Sequence	MATAQVELVQGGPRAPVGEKLELVLSNLQADVLELLLEFVYTGSLVIDSANAKTLLEA ASKFQFHTFCKVCVSFLEKQLTASNCLGVLAMAEAMQCSELYHMAKAFALQIFPEVAA QEEILSISKDDFIAYVSNDSLNTKAEELVYETVIKWIKKDPATRTQLQYAAELLAVVR LPFIHPSYLLNVVDNEELIKSSEACRDLVNEAKRYHMLPHARQEMQTPRTRPRVPAGV AEVIVLVGGRQMVGMTQRSLVAVTCWNPQNNKWYPLASLPFYDREFFSVVSAGDNIYL SGGMESGVTLADVWCYMSLLDNWNLVSRMTVPRCRHNSLVYDGKIYTLGGLGVAGNVD HVEVPAGVAEVIVLVGGRQMVGMTQRSLVAVTCWNPQNNKWYPLASLGGMESGVTLAD VWCYMSLLDNWNLVSRMTVPRCRHNSLVYDGKIYTLGGLGVAGNVDHVEAYEPTTNTW TLLPHMPCPVFRHGCVVIKKYIQSG		

Further analysis of the NOV23a protein yielded the following properties shown in Table 23B.

Table 23B. Protein Sequence Properties NOV23a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.2271 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space; 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV23a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 23C.

Table 23C. Geneseq Results for NOV23a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB40940	Human ORFX ORF704 polypeptide sequence SEQ ID NO:1408 - Homo sapiens, 335 aa. [WO200058473-A2, 05-OCT-2000]	19..351 4..334	317/333 (95%) 320/333 (95%)	e-180
AAM38711	Human polypeptide SEQ ID NO 1856 - Homo sapiens, 574 aa. [WO200153312-A1, 26-JUL-2001]	22..472 78..559	151/488 (30%) 222/488 (44%)	2e-61
AAB43090	Human ORFX ORF2854 polypeptide sequence SEQ ID NO:5708 - Homo sapiens, 506 aa. [WO200058473-A2, 05-OCT-2000]	22..468 9..487	150/491 (30%) 241/491 (48%)	3e-59
AAM38956	Human polypeptide SEQ ID NO 2101 - Homo sapiens, 587 aa. [WO200153312-A1, 26-JUL-2001]	22..468 90..568	149/491 (30%) 240/491 (48%)	1e-58
AAM94018	Human stomach cancer expressed polypeptide SEQ ID NO 106 - Homo sapiens, 568 aa. [WO200109317-A1, 08-FEB-2001]	25..470 76..553	148/490 (30%) 231/490 (46%)	3e-56

In a BLAST search of public sequence databases, the NOV23a protein was found to have homology to the proteins shown in the BLASTP data in Table 23D.

Table 23D. Public BLASTP Results for NOV23a				
Protein Accession Number	Protein/Organism/Length	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96CT2	HYPOTHETICAL 76.8 KDA PROTEIN - Homo sapiens (Human), 707 aa (fragment).	19..489 203..707	390/507 (76%) 406/507 (79%)	0.0
Q96PW7	KIAA1921 PROTEIN - Homo sapiens (Human), 545 aa (fragment).	19..489 41..545	390/507 (76%) 406/507 (79%)	0.0
Q96BF0	SIMILAR TO HYPOTHETICAL PROTEIN FLJ14106 - Homo sapiens (Human), 503 aa.	19..351 172..502	329/333 (98%) 330/333 (98%)	0.0

| Q9D5K3 | 4930429H24RIK PROTEIN - Mus musculus (Mouse), 484 aa. | 33..485
1..477 | 165/492 (33%)
248/492 (49%) | 2e-66 |
| Q9UH77 | Kelch-like protein 3 - Homo sapiens (Human), 587 aa. | 22..468
90..568 | 150/491 (30%)
241/491 (48%) | 1e-58 |

PFam analysis predicts that the NOV23a protein contains the domains shown in the Table 23E.

Table 23E. Domain Analysis of NOV23a			
Pfam Domain	NOV23a Match Region	Identities/ Similarities for the Matched Region	Expect Value
BTB: domain 1 of 1	4..79	24/143 (17%) 53/143 (37%)	3.7
Kelch: domain 1 of 4	223..272	9/50 (18%) 28/50 (56%)	0.94
Kelch: domain 2 of 4	275..320	11/47 (23%) 27/47 (57%)	0.016
Kelch: domain 3 of 4	322..396	14/75 (19%) 44/75 (59%)	3.3e-05
Kelch: domain 4 of 4	426..471	19/47 (40%) 35/47 (74%)	7.2e-10

Example 24.

The NOV24 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 24A.

Table 24A. NOV24 Sequence Analysis		
	SEQ ID NO: 81	4268 bp
NOV24a, CG57399-01 DNA Sequence	ATGACCTGGGACACAGCTCTCTGGACCTCAGTTTTTCTGATTGGGCTCCTTCCTACCC TTGGTTTCGCTAATTGCATCCTCCAGACTTCTGGTAAAATGTGTACTTTAAGAGGTAG ATACCCCCAGCCCCCACAACCACCTCTCTGCTTGTCTCCCCTAGTCCACCAGCTCCGA CCAGCAGACATCAAAGTGGTGGCCGCCCTGGGTAATGATGAAACCTTCCAGGAAAGTG GTGCAGGGCAGCTAAGTGAGCCTGACCCCAGGCAGTGGTCCTGGCCACAGGCCTGCTT GCCTGGGGTAAAAAAGGAAATGCAAGATGTGGTAGGTGAGAGAACGCCGAGCCGTCGC CGCAGCCTCCGCCGCCGAGAAGCCCTTGTTCCCGCTGCTGGGAAGGAGAGTCTGTGCC GACAAGATATTTTCATTTCCTTGTTGGAAATTATCAAGCATTTTCCTCCCTCCCCTCA GGACATCAACCTGGAGAAAGACTGGAAGCTGGTCACACTCTTCATTGGGGTCAACGAC TTGTGTCATTACTGTCCACTTGTTCAGGGCCCCGTTATAGACCTGGGTGGGATGGATA CCCTCCACTCCCTGCAGCTCCCAAGGGCTTTCGTCAACGTGGTGGAGGTCATGGAGCT GGCTAGCCTGTACCAGGGCCAAGGCGGGAAATGTGCCATGCTGGCAGCTCAGGAAGCC TGGAACAGCCTCCTGGCCTCCAGCAGGTACAGTGAGCAGGAGTCCTTCACCGTGGTTT TCCAGCCTTTCTTCTATGAGACCACCCCATCTGACCCCCGACTCCAGGATTCTACCAC GCTGGCCTGGCATCTCTGGAATAGGATGATGGAGCCAGCAGGAGAGAAAGATGAGCCA TTGAGTGTAAAACACGGGAGGCCAATGAAGTGTCCCTCTCAGGAGAGCCCCTATCTGT TCAGCTACAGAAACAGCAACTACCTGACCAGACTGCAGAAACCCCAAGACAAGCTTGT AAGAGAAGGAGCGGAAATCAGATGTCCTGACAAAGACCCCTCCGATACGGTTCCCACC TCAGTTCATAGGCTGAAGCCGGCTGACATCAACGTAATTGGAGCCCTGGGTGACTCTC TCACGGCAGGCAATGGGGCCGGGTCCACACCTGGGAACGTCTTGGACGTCTTGACTCA GTACCGAGGCCTGTCCTGGAGCGTCGGCGGAGATGAGAACATCGGCACCGTTACCACC

171

	CTGGCAGACATCCTCCGGGAATTCAACCCTTCCCTGAAGGGCTTCTCTGTTGGCACTG
	GGAAAGAAACCAGTCCTAATGCCTTCTTAAACCAGGCTGTGGCAGGAGGCCGAGCTGA
	GCAGGCCAGGAGGCTGGTGGACCTGATGAAGAATGACACGAGGATACACTTTCAGGAA
	GACTGGAAGATAATAACCCTGTTTATAGGCGGCAATGACCTCTGTGATTTCTGCAATG
	ATCTGGTACACTATTCTCCCCAGAACTTCACAGACAACATTGGAAAGGCCCTGGACAT
	CCTCCATGCTGAGTCTCAGGTTCCTCGGGCATTTGTGAACCTGGTGACGGTGCTTGAG
	ATCGTCAACCTGAGGGAGCTGTACCAGGAGAAAAAAGTCTACTGCCCAAGGATGATCC
	TCAGGTCACTGTGTCCCTGTGTCCTGAAGTTTGATGATAACTCAACAGAACTTGCTAC
	CCTCATCGAATTCAACAAGAAGTTTCAGGAGAAGACCCACCAACTGATTGAGAGTGGG
	CGATATGACACAAGGGAAGATTTTACTGTGGTTGTGCAGCCGTTCTTTGAAAACGTGG
	ACATGCCAAAGACCCAGGAAGGATTGCCTGACAACTCTTTCTTCGCTCCTGACTGTTT
	CCACTTCAGCAGCAAGTCTCACTCCCGAGCAGCCAGTGCTCTCTGGAACAATATGCTG
	GAGCCTGTTGGCCAGAAGACGACTCGTCATAAGTTTGAAAACAAGATCAATATCACAT
	GTCCGTCACAGGTCCAGCCGTTTCTGAGGACCTACAAGAACAGCATGCAGGGTCATGG
	GACCTGGCTGCCATGCAGGGACAGAGCCCCTTCTGCCTTGCACCCTACCTCAGTGCAT
	GCCCTGAGACCTGCAGACATCCAAGTTGTGGCTGCTCTGGGGGATTCTCTGACCGCTG
	GCAATGGAATTGGCTCCAAACCAGACGACCTCCCCGATGTCACCACACAGTATCGGGG
	ACTGTCATACAGTGCAGGAGGGGACGGCTCCCTGGAGAATGTGACCACCTTACCTAGT
	TCTATCCTTCGGGAGTTTAACAGAAACCTCACAGGCTACGCCGTGGGGCACGGGTGATG
	CCAATGACACGAATGCATTCCTCAATCAAGCTGTTCCCGGAGCAAAGGCTAGGGATCT
	TATGAGCCAAGTCCAAACTCTGATGCAGAAGATGAAAGATGATCATAGAGTAAATTTC
	CATGAAGACTGGAAGGTCATCACAGTGCTGATCGGAGGCAGCGATTTATGTGACTACT
	GCACAGATTCGAATCTGTATTCTGCAGCCAACTTTGTTCACCATCTCCGCAATGCCTT
	GGACGTCCTGCATAGAGAGGTGCCCAGAGTCCTGGTCAACCTCGTGGACTTCCTGAAC
	CCCACTATCATGCGGCAGGTGTTCCTGGGAAACCCAGACAAGTGCCCAGTGCAGCAGG
	CCAGCGTTTTGTGTAACTGCGTTCTGACCCTGCGGGAGAACTCCCAAGAGCTAGCCAG
	GCTGGAGGCCTTCAGCCGAGCCTACCAGAGCAGCATGCGCGAGCTGGTGGGGTCAGGC
	CGCTATGACACGCAGGAGGACTTCTCTGTGGTGCTGCAGCCCTTCTTCCAGAACATCC
	AGCTCCCTGTCCTGCAGGATGGGCTCCCAGATACGTCCTTCTTTGCCCCAGACTGCAT
	CCACCCAAATCAGAAATTCCACTCCCAGCTGGCCAGAGCCCTTTGGACCAATATGCTT
	GAACCACTTGGAAGCAAAACAGAGACCCTGGACCTGAGAGCAGAGATGCCCATCACCT
	GTCCCACTCAGAATGAGCCCTTCCTGAGAACCCCTCGGAATAGTAACTACACGTACCC
	CATCAAGCCAGCCATTGAGAACTGGGGCAGTGACTTCCTGTGTACAGAGTGGAAGGCT
	TCCAATAGTGTTCCAACCTCTGTCCACCAGCTCCGACCAGCAGACATCAAAGTGGTGG
	CCGCCCTGGGTGACTCTCTGACTACAGCAGTGGGAGCTCGACCAAACAACTCCAGTGA
	CCTACCCACATCTTGGAGGGGACTCTCTTGGAGCATTGGAGGGGATGGGAACTTGGAG
	ACTCACACCACACTGCCCAGTATTCTGAAGAAGTTCAACCCTTACCTCCTTGGCTTCT
	CTACCAGCACCTGGGAGGGGACAGCAGGACTAAATGTGGCAGCGGAAGGGGCCAGAGC
	TAGGAGGGACATGCCAGCCCAGGCCTGGGACCTGGTAGAGCGAATGAAAAACAGCCCC
	ATACACTTTCAGGAAGACTGGAAGATAATAACCCTGTTTATAGGCGGCAATGACCTCT
	GTGATTTCTGCAATGATCTGGTAGGTGAATATGTTCAGCACATCCAACAGGCCCTGGA
	CATCCTCTCTGAGGAGCTCCCAAGGGCTTTCGTCAACGTGGTGGAGGTCATGGAGCTG
	GCTAGCCTGTACCAGGGCCAAGGCGGGAAATGTGCCATGCTGGCAGCTCAGAACAACT
	GCACTTGCCTCAGACACTCGCAAAGCTCCCTGGAGAAGCAAGAACTGAAGAAAGTGAA
	CTGGAACCTCCAGCATGGCATCTCCAGTTTCTCCTACTGGCACCAATACACACAGCGT
	GAGGACTTTGCGGTTGTGGTGCAGCCTTTCTTCCAAAACACACTCACCCCACTGAACA
	GAGGGGACACTGACCTCACCTTCTTCTCCGAGGACTGTTTTCACTTCTCAGACCGCGG
	GCATGCCGAGATGGCCATCGCACTCTGGAACAACATGCTGGAACCAGTGGGCCGCAAG
	ACTACCTCCAACAACTTCACCCACAGCCGAGCCAAACTCAAGTGCCCCTCTCCTGTGA
	GTCCTTACCTCTACACCCTGCGGAACAGCCGATTGCTCCCAGACCAGGCTGAAGAAGC
	CCCCGAGGTGCTCTACTGGGCTGTCCCAGTGGCAGCGGGAGTCGGCCTTGTGGTGGGC
	ATCATCGGGACAGTGGTCTGGAGGTGCAGGAGAGGTGGCCGGAGGGAAGATCCTCCAA
	TGAGCCTGCGCACTGTGGCCCTC**TAG**GCCCGGGG

ORF Start: ATG at 1		ORF Stop: TAG at 4258
SEQ ID NO: 82	1419 aa	MW at 158435.1kD

| NOV24a,
CG57399-01 Protein
Sequence | MTWDTALWTSVFLIGLLPTLGFANCILQTSGKMCTLRGRYPQPPQPPLCLSPLVHQLR
PADIKVVAALGNDETFQESGAGQLSEPDPRQWSWPQACLPGVKKEMQDVVGERTPSRR
RSLRRREALVPAAGKESLCRQDIFISLLEIIKHFPPSPQDINLEKDWKLVTLFIGVND
LCHYCPLVQGPVIDLGGMDTLHSLQLPRAFVNVVEVMELASLYQGQGGKCAMLAAQEA
WNSLLASSRYSEQESFTVVFQPFFYETTPSDPRLQDSTTLAWHLWNRMMEPAGEKDEP
LSVKHGRPMKCPSQESPYLFSYRNSNYLTRLQKPQDKLVREGAEIRCPDKDPSDTVPT
SVHRLKPADINVIGALGDSLTAGNGAGSTPGNVLDVLTQYRGLSWSVGGDENIGTVTT
LADILREFNPSLKGFSVGTGKETSPNAFLNQAVAGGRAEQARRLVDLMKNDTRIHFQE
DWKIITLFIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAESQVPRAFVNLVTVLE
IVNLRELYQEKKVYCPRMILRSLCPCVLKFDDNSTELATLIEFNKKFQEKTHQLIESG
RYDTREDFTVVVQPFFENVDMPKTQEGLPDNSFFAPDCFHFSSKSHSRAASALWNNML
EPVGQKTTRHKFENKINITCPSQVQPFLRTYKNSMQGHGTWLPCRDRAPSALHPTSVH
ALRPADIQVVAALGDSLTAGNGIGSKPDDLPDVTTQYRGLSYSAGGDGSLENVTTLPS
SILREFNRNLTGYAVGTGDANDTNAFLNQAVPGAKARDLMSQVQTLMQKMKDDHRVNF
HEDWKVITVLIGGSDLCDYCTDSNLYSAANFVHHLRNALDVLHREVPRVLVNLVDFLN
PTIMRQVFLGNPDKCPVQQASVLCNCVLTLRENSQELARLEAFSRAYQSSMRELVGSG
RYDTQEDFSVVLQPFFQNIQLPVLQDGLPDTSFFAPDCIHPNQKFHSQLARALWTNML
EPLGSKTETLDLRAEMPITCPTQNEPFLRTPRNSNYTYPIKPAIENWGSDFLCTEWKA
SNSVPTSVHQLRPADIKVVAALGDSLTTAVGARPNNSSDLPTSWRGLSWSIGGDGNLE |

172

	THTTLPSILKKFNPYLLGFSTSTWEGTAGLNVAAEGARARRDMPAQAWDLVERMKNSP IHFQEDWKIITLFIGGNDLCDFCNDLVGEYVQHIQQALDILSEELPRAFVNVVEVMEL ASLYQGQGGKCAMLAAQNNCTCLRHSQSSLEKQELKKVNWNLQHGISSFSYWHQYTQR EDFAVVVQPFFQNTLTPLNRGDTDLTFFSEDCFHFSDRGHAEMAIALWNNMLEPVGRK TTSNNFTHSRAKLKCPSPVSPYLYTLRNSRLLPDQAEEAPEVLYWAVPVAAGVGLVVG IIGTVVWRCRRGGRREDPPMSLRTVAL	
	SEQ ID NO: 83	1624 bp
NOV24b, CG57399-02 DNA Sequence	GCCGGCTGACATCAATGTAATTGGAGCCCTGGGTGACTCTCTCACGGCAGGCAATGGG GCCGGGTCCACACCTGGGAACGTCTTGGACGTCTTGACTCAGTACCGAGGCCTGTCCT GGAGCGTCGGCGGAGATGAGAACATCGGCACCGTTACCACCCTGGCGAACATCCTCCG GGAATTCAACCCTTCCCTGAAGGGCTTCTCTGTTGGCACTGGGAAAGAAACCAGTCCT AATGCCTTCTTAAACCAGGCTGTGGCAGGAGGCCGAGCTGAGGATCTACCTGTCCAGG CCAGGAGGCTGGTGGACCTGATGAAGAATGACACGAGGATACACTTTCAGGAAGACTG GAAGATAATAACCCTGTTTATAGGCGGCAATGACCTCTGTGATTTCTGCAATGATCTG GTCCACTATTCTCCCCAGAACTTCACAGACAACATTGGAAAGGCCCTGGACATCCTCC ATGCTGAGGTTCCTCGGGCATTTGTGAACCTGGTGACGGTGCTTGAGATCGTCAACCT GAGGGAGCTGTACCAGGAGAAAAAAGTCTACTGCCCAAGGATGATCCTCAGGTCTCTG TGTCCCTGTGTCCTGAAGTTTGATGATAACTCAACAGAACTTGCTACCCTCATCGAAT TCAACAAGAAGTTTCAGGAGAAGACCCACCAACTGATTGAGAGTGGGCGATATGACAC AAGGGAAGATTTTACTGTGGTTGTGCAGCCGTTCTTTGAAAACGTGGACATGCCAAAG ACCTCGGAAGGATTGCCTGACAACTCTTTCTTCGCTCCTGACTGTTTCCACTTCAGCA GCAAGTCTCACTCCCGAGCAGCCAGTGCTCTCTGGAACAATATGCTGGAGCCTGTTGG CCAGAAGACGACTCGTCATAAGTTTGAAAACAAGATCAATATCACATGTCCGAACCAG GTCCAGCCGTTTCTGAGGACCTACAAGAACAGCATGCAGGGTCATGGGACCTGGCTGC CATGCAGGGACAGAGCCCCTTCTGCCTTGCACCCTACCTCAGTGCATGCCCTGAGACC TGCAGACATCCAAGTTGTGGCTGCTCTGGGGGATTCTCTGACCGCTGGCAATGGAATT GGCTCCAAACCAGACGACCTCCCCGATGTCACCACACAGTATCGGGGACTGTCATACA GAGAAAGTAAACCAGGGTTCTTATCAGACTCCTGGGTCAGCAAATCCAACAGGAAATG CACCAGAAAAGCACCAAATCCCTGAATCTTCACCTCCCCGCTTGCATGTATACGTGTA CACGTGGTGTTCCTACGTCTCTGTTTACTGTCTTTATGTGTTTATTCATGTTGTCTTG TAGTCACACAGCTGCCTTTACATATATGTACACATCTGCACAGAAAACCTCTGAAACC CATCGCACACTTCGAGAGGCCATAACCAAGACACAATCACAATCAGCCATGTCTTGAA AGATTAGCAATTCGACAAGAGGAAAGGGTGAGAAAGGGCATCCCGAACACGGAAGTGG AGAAGCTCAGGGTGTGTCAGGCGAGCGGTTGCGTGTAGATATTCTCAAGTTTCTTTCT CTCCTAATAAAGTTCTCATTCCTGTAGGCTTCAAAGTAAGTGGCGAGTAGCTCAGAAT	
	ORF Start: ATG at 311	**ORF Stop: TGA at 1241**
	SEQ ID NO: 84	310 aa	MW at 35240.6kD
NOV24b, CG57399-02 Protein Sequence	MKNDTRIHFQEDWKIITLFIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAEVPRA FVNLVTVLEIVNLRELYQEKKVYCPRMILRSLCPCVLKFDDNSTELATLIEFNKKFQE KTHQLIESGRYDTREDFTVVVQPFFENVDMPKTSEGLPDNSFFAPDCFHFSSKSHSRA ASALWNNMLEPVGQKTTRHKFENKINITCPNQVQPFLRTYKNSMQGHGTWLPCRDRAP SALHPTSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLPDVTTQYRGLSYRESKPGF LSDSWVSKSNRKCTRKAPNP		
	SEQ ID NO: 85	4425 bp
NOV24c, CG57399-03 DNA Sequence	CTGGAGCATTCTGGCATGGGGCTGCGGCCAGGCATTTTCCTCCTGGAGCTGCTGCTGC TTCTGGGGCAAGGTACCCCTCAGATCCATACCTCTCCTAGAAAGAGTACATTGGAAGG GCAGCTATGGCCAGAGACAGTTCACTCTCTGAAGCCTTCTGATATTAAATTTGTGGCA GCCATTGGCAATCTGGAAATTGTGCCAGACCCAGGGACGGGCGATCTGGAGAAGCAAG ACGAAAGGCCACAGCAGGTGTGCATGGGAGTGATGACAGTCCTTTCAGACATCATCAG ATATTTCAGTCCTTCTGTTCCAATGCCTGTGTGCCACACTGGAAAGAGAGTCATACCC CACGATGGTGCTGAGGACTTGTGGATTCAGGCTCAAGAACTGGTGAGAAACATGAAAG AGAACCAACTTGACTTTCAATTTGACTGGAAGCTCATCAATGTGTTCTTCAGTAATGC AAGCCAGTGTTACCTGTGCCCCTCTGCTCAACAGAATGGGCTTGCGGCGGGCGGCGTG GATGAGCTGATGGGGGTGCTGGACTACCTGCAGCAGGAGGTGCCCAGAGCATTTGTAA ACCTGGTGGACCTCTCTGAGGTTGCAGAGGTCTCTCGTCAGTATCACGGCACTTGGCT CAGCCCTGCACCAGAGCCCTGTAATTGCTCAGAGGAGACCACCCGGCTGGCCAAGGTG GTGATGCAGTGGTCTTATCAGGAAGCCTGGAACAGCCTCCTGGCCTCCAGCAGGTACA GTGAGCAGGAGTCCTTCACCGTGGTTTTCCAGCCTTTCTTCTATGAGACCACCCCATC TGACCCCCGACTCCAGGATTCTACCACGCTGGCCTGGCATCTCTGGAATAGGATGATG GAGCCAGCAGGAGAGAAAGATGAGCCATTGAGTGTAAAACACGGGAGGCCAATGAAGT GTCCCTCTCAGGAGAGCCCCTATCTGTTCAGCTACAGAAACAGCAACTACCTGACCAG ACTGCAGAAACCCCAAGACAAGCTTGAGGTAAGAGAAGGAGCGGAAATCAGATGTCCT GACAAAGACCCCTCCATACGGTTCCCCACCTCAGTTCATAGGCTGAAGCCGGCTGACA TCAACGTAATTGGAGCCCTGGGTGACTCTCTCACGGCAGGCAATGGGGCCGGGTCCAC ACCTGGGAACGTCTTGGACGTCTTGACTCAGTACCGAGGCCTGTCCTGGAGCGTCGGC GGAGATGAGAACATCGGCACCGTTACCACCCTGGCGGACATCCTCCGGGAATTCAACC CTTCCCTGAAGGGCTTCTCTGTTGGCACTGGGAAAGAAACCAGTCCTAATGCCTTCTT AAACCAGGCTGTGGCAGGAGGCCGAGCTGAGGATCTACCTGTCCAGGCCAGGAGGCTGGTGGACCTGATG AAGAATGACACGAGGATACACTTTCAGGAAGACTGGAAGATAATAACCCTGTTTATAG GCGGCAATGACCTCTGTGATTTCTGCAATGATCTGGTACACTATTCTCCCCAGAACTT	

	CACAGACAACATTGGAAAGGCCCTGGACATCCTCCATGCTGAGGTTCCTCGGGCATTT GTGAACCTGGTGACGGTGCTTGAGATCGTCAACCTGAGGGAGCTGTACCAGGAGAAAA AAGTCTACTGCCCAAGGATGATCCTCAGGTCACTGTGTCCCTGTGTCCTGAAGTTTGA TGATAACTCAACAGAACTTGCTACCCTCATCGAATTCAACAAGAAGTTTCAGGAGAAG ACCCACCAACTGATTGAGAGTGGGCGATATGACACAAGGGAAGATTTTACTGTGGTTG TGCCAGCCGTTCTTTGAAAACGTGGACATGCCAAAGACCCAGGAAGGATTGCCTGACAA CTCTTTCTTCGCTCCTGACTGTTTCCACTTCAGCAGCAAGTCTCACTCCCGAGCAGCC AGTGCTCTCTGGAACAATATGCTGGAGCCTGTTGGCCAGAAGACGACTCGTCATAAGT TTGAAAACAAGATCAATATCACATGTCCGAACCAGGTAGAGTGGCCGTTTCTGAGGAC CTACAAGAACAGCATGCAGGGTCATGGGACCTGGCTGCCATGCAGGGACAGAGCCCCT TCTGCCTTGCACCCTACCTCAGTGCATGCCCTGAGACCTGCAGACATCCAAGTTGTGG CTGCTCTGGGGGATTCTCTGACCGCTGGCAATGGAATTGGCTCCAAACCAGACGACCT CCCCGATGTCACCACACAGTATCGGGGACTGTCATACAGTGCAGGAGGGGACGGCTCC CTGGAGAATGTGACCACCTTACCTGATATCCTTCGGGAGTTTAACAGAAACCTCACAG GCTACGCCGTGGGCACGGGTGATGCCAATGACACGAATGCATTCCTCAATCAAGCTGT TCCCGGAGCAAAGGCTAGGGATCTTATGAGCCAAGTCCAAACTCTGATGCAGAAGATG AAAGATGATCATAGAGTAAATTTCCATGAAGACTGGAAGGTCATCACAGTGCTGATCG GAGGCAGCGATTTATGTGACTACTGCACAGATTCGAATCTGTATTCTGCAGCCAACTT TGTTCACCATCTCCGCAATGCCTTGGACGTCCTGCATAGAGAGGTGCCCAGAGTCCTG GTCAACCTCGTGGACTTCCTGAACCCCACTATCATGCGGCAGGTGTTCCTGGGAAACC CAGACAAGTGCCCAGTGCAGCAGGCCAGCGTTTTGTGTAACTGCGTTCTGACCCTGCG GGAGAACTCCCAAGAGCTAGCCAGGCTGGAGGCCTTCAGCCGAGCCTACCAGAGCAGC ATGCGCGAGCTGGTGGGGTCAGGCCGCTATGACACGCAGGAGGACTTCTCTGTGGTGC TGCAGCCCTTCTTCCAGAACATCCAGCTCCCTGTCCTGCAGGATGGGCTCCCAGATAC GTCCTTCTTTGCCCCAGACTGCATCCACCCAAATCAGAAATTCCACTCCCAGCTGGCC AGAGCCCTTTGGACCAATATGCTTGAACCACTTGGAAGCAAAACAGAGACCCTGGACC TGAGAGCAGAGATGCCCATCACCTGTCCCACTCAGAATGAGCCCTTCCTGAGAACCCC TCGGAATAGTAACTACACGTACCCCATCAAGCCAGCCATTGAGAACTGGGGCAGTGAC TTCCTGTGTACAGAGTGGAAGGCTTCCAATAGTGTTCCAACCTCTGTCCACCAGCTCC GACCAGCAGACATCAAAGTGGTGGCCGCCCTGGGTGACTCTCTGACTGTGGCAGTGGG AGCTCGACCAAACAACTCCAGTGACCTACCCACATCTTGGAGGGGACTCTCTTGGAGC ATTGGAGGGGATGGGAACTTGGAGACTCACACCACACTGCCCGACATTCTGAAGAAGT TCAACCCTTACCTCCTTGGCTTCTCTACCAGCACCTGGGAGGGGACAGCAGGACTAAA TGTGGCAGCGGAAGGGGCCAGAGCTAGGGACATGCCAGCCCAGGCCTGGGACCTGGTA GAGCGAATGAAAAACAGCCCCCAGGACATCAACCTGGAGAAAGACTGGAAGCTGGTCA CACTCTTCATTGGGGTCAACGACTTGTGTCATTACTGTGAGAATCCGGTAGGCGAATA TGTTCAGCACATCCAACAGGCCCTGGACATCCTCTCTGAGGAGCTCCCAAGGGCTTTC GTCAACGTGGTGGAGGTCATGGAGCTGGCTAGCCTGTACCAGGGCAAGGCGGGAAAT GTGCCATGCTGGCAGCTCAGAACAACTGCACTTGCCTCAGACACTCGCAAAGCTCCCT GGAGAAGCAAGAACTGAAGAAAGTGAACTGGAACCTCCAGCATGGCATCTCCAGTTTC TCCTACTGGCACCAATACACACAGCGTGAGGACTTTGCGGTTGTGGTGCAGCCTTTCT TCCAAAACACACTCACCCCACTGAACAGAGGGGACACTGACCTCACCTTCTTCTCCGA GGACTGTTTCACTTCTCAGACCGCGGGCATGCCGAGATGGCCATCGCACTCTGGAAC AACATGCTGGAACCAGTGGGCCGCAAGACTACCTCCAACAACTTCACCCACAGCCGAG CCAAACTCAAGTGCCCCTCTCCTGAGAGCCCTTACCTCTACACCCTGCGGAACAGCCG ATTGCTCCCAGACCAGGCTGAAGAAGCCCCCGAGGTGCTCTACTGGGCTGTCCCAGTG GCAGCGGGAGTCGGCCTTGTGGTGGGCATCATCGGGACAGTGGTCTGGAGGTGCAGGA GAGGTGGCCGGAGGGAAGATCCTCCAATGAGCCTGCGCACTGTGGCCCTC**TAG**GCCCG <u>GGGGTGGGTCCTCACCCTAAACTCCCTATAGCCACTCTTCTTCACCGCCCTCTGCCCCA</u> <u>GCCACTCCCGGCCACCAGGACATGCTTCAATGCCTGGTGCCATAGGAAGCCCAGGGGA</u> <u>CAGTCACAACTTCTTGG</u>
	ORF Start: ATG at 16 **ORF Stop: TAG at 4285**
	SEQ ID NO: 86 **1423 aa** **MW at 159352.7kD**
NOV24c, **CG57399-03 Protein** **Sequence**	MGLRPGIFLLELLLLLGQGTPQIHTSPRKSTLEGQLWPETVHSLKPSDIKFVAAIGNL EIVPDPGTGDLEKQDERPQQVCMGVMTVLSDIIRYFSPSVPMPVCHTGKRVIPHDGAE DLWIQAQELVRNMKENQLDFQFDWKLINVFFSNASQCYLCPSAQQNGLAAGGVDELMG VLDYLQQEVPRAFVNLVDLSEVAEVSRQYHGTWLSPAPEPCNCSEETTRLAKVVMQWS YQEAWNSLLASSRYSEQESFTVVFQPFFYETTPSDPRLQDSTTLAWHLWNRMMEPAGE KDEPLSVKHGRPMKCPSQESPYLFSYRNSNYLTRLQKPQDKLEVREGAEIRCPDKDPS DTVPTSVHRLKPADINVIGALGDSLTAGNGAGSTPGNVLDVLTQYRGLSWSVGGDENI GTVTTLADILREFNPSLKGFSVGTGKETSPNAFLNQAVAGGRAEQARRLVDLMKNDTR IHFQEDWKIITLFIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAEVPRAFVNLVT VLEIVNLRELYQEKKVYCPRMILRSLCPCVLKFDDNSTELATLIEFNKKFQEKTHQLI ESGRYDTREDFTVVVQPFFENVDMPKTQEGLPDNSFFAPDCFHFSSKSHSRAASALWN NMLEPVGQKTTRHKFENKINITCPNQVEWPFLRTYKNSMQGHGTWLPCRDRAPSALHP TSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLPDVTTQYRGLSYSAGGDGSLENVT TLPDILREFNRNLTGYAVGTGDANDTNAFLNQAVPGAKARDLMSQVQTLMQKMKDDHR VNFHEDWKVITVLIGGSDLCDYCTDSNLYSAANFVHHLRNALDVLHREVPRVLVNLVD FLNPTIMRQVFLGNPDKCPVQQASVLCNCVLTLRENSQELARLEAFSRAYQSSMRELV GSGRYDTQEDFSVVLQPFFQNIQLPVLQDGLPDTSFFAPDCIHPNQKFHSQLARALWT NMLEPLGSKTETLDLRAEMPITCPTQNEPFLRTPRNSNYTYPIKPAIENWGSDFLCTE WKASNSVPTSVHQLRPADIKVVAALGDSLTVAVGARPNNSSDLPTSWRGLSWSIGGDG NLETHTTLPDILKKFNPYLLGFSTSTWEGTAGLNVAAEGARARDMPAQAWDLVERMKN SPQDINLEKDWKLVTLFIGVNDLCHYCENPVGEYVQHIQQALDILSEELPRAFVNVVE

174

	VMELASLYQGQGGKCAMLAAQNNCTCLRHSQSSLEKQELKKVNWNLQHGISSFSYWHQ YTQREDFAVVVQPFFQNTLTPLNRGDTDLTFFSEDCFHFSDRGHAEMAIALWNNMLEP VGRKTTSNNFTHSRAKLKCPSPESPYLYTLRNSRLLPDQAEEAPEVLYWAVPVAAGVG LVVGIIGTVVWRCRRGGRREDPPMSLRTVAL

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 24B.

Table 24B. Comparison of NOV24a against NOV24b through NOV24c.

Protein Sequence	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV24b	454..748 1..293	283/295 (95%) 285/295 (95%)
NOV24c	27..1419 23..1423	1211/1426 (84%) 1261/1426 (87%)

Further analysis of the NOV24a protein yielded the following properties shown in

5   Table 24C.

Table 24C. Protein Sequence Properties NOV24a

PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1080 probability located in nucleus
SignalP analysis:	Likely cleavage site between residues 24 and 25

A search of the NOV24a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 24D.

Table 24D. Geneseq Results for NOV24a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW30751	Rat phospholipase-B/lipase - Rattus rattus, 1450 aa. [JP09248190-A, 22-SEP-1997]	50..1403 60..1447	911/1404 (64%) 1077/1404 (75%)	0.0
ABB11053	Human phospholipase B	985..1203 45..267	205/224 (91%) 213/224 (94%)	e-117

	Homo sapiens, 267 aa. [WO200157188-A2, 09-AUG-2001]			
AAM25824	Human protein sequence SEQ ID NO:1339 - Homo sapiens, 267 aa. [WO200153455-A2, 26-JUL-2001]	985..1203 45..267	205/224 (91%) 213/224 (94%)	e-117
AAM95420	Human reproductive system related antigen SEQ ID NO: 4078 - Homo sapiens, 148 aa. [WO200155320-A2, 02-AUG-2001]	979..1106 4..133	110/130 (84%) 117/130 (89%)	3e-56
ABB11237	Human phospholipase homologue, SEQ ID NO:1607 - Homo sapiens, 132 aa. [WO200157188-A2, 09-AUG-2001]	393..478 43..132	84/90 (93%) 86/90 (95%)	3e-40

In a BLAST search of public sequence databases, the NOV24a protein was found to have homology to the proteins shown in the BLASTP data in Table 24E.

Table 24E. Public BLASTP Results for NOV24a				
Protein Accession Number	Protein/Organism/Length	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q05017	Phospholipase ADRAB-B precursor (EC 3.1.-.-) - Oryctolagus cuniculus (Rabbit), 1458 aa.	6..1416 2..1456	1042/1466 (71%) 1179/1466 (80%)	0.0
O70320	PHOSPHOLIPASE B - Cavia porcellus (Guinea pig), 1463 aa.	7..1414 3..1458	965/1474 (65%) 1135/1474 (76%)	0.0
O54728	PHOSPHOLIPASE B - Rattus norvegicus (Rat), 1450 aa.	50..1403 60..1447	911/1404 (64%) 1077/1404 (75%)	0.0
Q96DP9	CDNA FLJ30866 FIS, CLONE FEBRA2004110, HIGHLY SIMILAR TO PHOSPHOLIPASE ADRAB-B PRECURSOR (EC 3.1.-.-) - Homo sapiens (Human), 270 aa.	454..714 1..259	257/261 (98%) 258/261 (98%)	e-151
Q9N2Z4	HYPOTHETICAL 41.4 KDA PROTEIN - Caenorhabditis elegans, 377 aa.	343..673 37..369	130/343 (37%) 202/343 (57%)	1e-59

176

PFam analysis predicts that the NOV24a protein contains the domains shown in the Table 24F.

Table 24F. Domain Analysis of NOV24a			
Pfam Domain	NOV24a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Lipase_GDSL: domain 1 of 3	360..484	54/147 (37%) 116/147 (79%)	4.8e-42
Lipase_GDSL: domain 2 of 3	705..834	57/147 (39%) 116/147 (79%)	4.5e-44
SecA_protein: domain 1 of 1	834..851	10/20 (50%) 17/20 (85%)	4.9
Vitellogenin_N: domain 1 of 1	1107..1124	8/18 (44%) 17/18 (94%)	3.8
Lipase_GDSL: domain 3 of 3	1062..1185	48/147 (33%) 114/147 (78%)	6.3e-37

Example 25.

5   The NOV25 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 25A.

Table 25A. NOV25 Sequence Analysis		
	SEQ ID NO: 87	1348 bp
NOV25a, CG59311-01 DNA Sequence	CTGGGTCGCCCCTGTTCTACCCAGATTGGGATGGCAGCGACGCTGATCCTGGAGCCCG CGGGCCGCTGCTGCTGGGACGAGCCGCTGCGCATCGCAGTGCGCGGCCTGGCCCCGGA GCAGCCAGTCACGCTGCGCACGTCCCTGCGCGACGAAGAGGGCGCGCTCTTCCGGGCC CACGCGCGCTACCGTGCCGACGCCCGCGACGAGCTGGACCTGGAGCGCGCGCCCGCGC TGGGGAGGCAGCTTCGCGGGGCTCCAGCCCATGGGGCTGCTGTGGGCGTTGGAGCCCGA GAAAGCCTTGGTGCGGCTGGTGAAGCGCGACGTGCGGACGCCCTTCGCCGTGGAGCTG GAAGTGCTGGACGGCCACGACACCGAGCCCGGGCGGCTGCTGTGCCTGGCGCAGAACA AGCGCGACTTTCTCCGGCCGGGGGTGCGGCGCGAGCCGGTGCGCGCGGGCCCGGTGCG CGCCGCGCTCTTCCTGCCGCCGGATAGGGGGCCCTTTCCTGGGATCATTGATCTGTTT GGGAGCAGCAGGGGCCTTTGTGAATACAGGGCCAGCCTCCTGGCCGGACATGGTTTTG CTGTGCTTGCCCTGGCTTATTTCAGATTTGAAGACCTCCCCGAAGATCTGAATGATGT ACATCTGGAGTACTTTGAAGAAGCCGTGGACTTTATGCTGCAGCATCCAAAGGTGAAA GGTCCTAGTATTGCGCTTCTTGGATTTTCCAAAGGAGGTGACCTGTGTCTCTCAATGG CTTCTTTCTTGAAGGGCATCACAGCCACTGTACTTATCAATGCCTGTGTAGCCAACAC AGTAGCTCCTCTACATTACAAGGATATGATTATTCCTAAACTTGTCGATGATCTAGGA AAAGTAAAAATCACTAAGTCAGGATTTCTCACTTTTATGGACACTTGGAGCAATCCAC TGGAGGAACACAATCACCAAAGTCTTGTTCCATTGGAAAAGGCGCAGGTGCCCTTCTT GTTTATTGTTGGCATGGATGATCAAAGCTGGAAGAGTGAATTCTATGCTCAGATAGCC TCTGAAAGGCTACAAGCTCATGGGAAAGAAAGACCCCAGATAATCTGTTACCCAGAAA CTGGTCACTGTATTGACCCACCTTATTTTCCTCCTTCTAGAGCTTCTGTGCACGCTGT TTTGGGTGAGGCAATATTCTATGGAGGTGAGCCAAAGGCTCACTCAAAGGCACAGGTA GATGCCTGGCAGCAAATTCAAACTTTCTTCCATAAACATCTCAATGGTAAAAAATCTG TCAAGCACAGCAAAATATAACATTGTAGCCACAGACCAGATACCATTAATAAAAATCC TATTCATACAACTT
	ORF Start: ATG at 31	ORF Stop: TAA at 1294

	SEQ ID NO: 88	421 aa	MW at 46815.4kD
NOV25a, CG59311-01 Protein Sequence	MAATLILEPAGRCCWDEPLRIAVRGLAPEQPVTLRTSLRDEEGALFRAHARYRADARD ELDLERAPALGGSFAGLQPMGLLWALEPEKALVRLVKRDVRTPFAVELEVLDGHDTEP GRLLCLAQNKRDFLRPGVRREPVRAGPVRAALFLPPDRGPFPGIIDLFGSSRGLCEYR ASLLAGHGFAVLALAYFRFEDLPEDLNDVHLEYFEEAVDFMLQHPKVKGPSIALLGFS KGGDLCLSMASFLKGITATVLINACVANTVAPLHYKDMIIPKLVDDLGKVKITKSGFL TFMDTWSNPLEEHNHQSLVPLEKAQVPFLFIVGMDDQSWKSEFYAQIASERLQAHGKE RPQIICYPETGHCIDPPYFPPSRASVHAVLGEAIFYGGEPKAHSKAQVDAWQQIQTFF HKHLNGKKSVKHSKI		
	SEQ ID NO: 89	1021 bp	
NOV25b, CG59311-02 DNA Sequence	AGATTGGGATGGCAGCGACGCTGATCCTGGAGCCCGCGGGCCGCTGCTGCTGGGACGA GCCGCTGCGCATCGCAGTGCGCGGCCTGGCCCCGGAGCAGCCAGTCACGCTGCGCACG TCCCTGCGCGACGAAGAGGGCGCGCTCTTCCGGGCCCACGCGCGCTACCGTGCCGACG CCTCTAATCCCGGCACTTTGGGGGGCCAAGGCAGGGGGCCCTTTCCTGGGATCATTGA TCTGTTTGGGAGCAGCAGGGGCCTTTGTGAATACAGGGCCAGCCTCCTGGCCGGACAT GGTTTTGCTGTGCTTGCCCTGGCTTATTTCAGATTTGAAGACCTCCCCGAAGATCTGA ATGATGTACATCTGGAGTACTTTGAAGAAGCCGTGGACTTTATGCTGCAGCATCCAAA GGTGAAAGGTCCTAGTATTGCGCTTCTTGGATTTTCCAAAGGAGGTGACCTGTGTCTC TCAATGGCTTCTTTCTTGAAGGGCATCACAGCCACTGTACTTATCAATGCCTGTGTAG CCAACACAGTAGCTCCTCTACATTACAAGGATATGATTATTCCTAAACTTGTCGATGA TCTAGGAAAAGTAAAAATCACTAAGTCAGGATTTCTCACTTTTATGGACACTTGGAGC AATCCACTGGAGGAACACAATCACCAAAGTCTTGTTCCATTGGAAAAGGCGCAGGTGC CCTTCTTGTTTATTGTTGGCATGGATGATCAAAGCTGGAAGAGTGAATTCTATGCTCA GATAGCCTCTGAAAGGCTACAAGCTCATGGGAAAGAAAGACCCCAGATAATCTGTTAC CCAGAAACTGGTCACTGTATTGACCCACCTTATTTTCCTCCTTCTAGAGCTTCTGTGC ACGCTGTTTTGGGTGAGGCAATATTCTATGGAGGTGAGCCAAAGGCTCACTCAAAGGC ACAGGTAGATGCCTGGCAGCAAATTCAAACTTTCTTCCATAAACATCTCAATGGTAAA AAATCTGTCAAGCACAGCAAAATATAACATTGTAG		
	ORF Start: ATG at 9	ORF Stop: TAA at 1011	
	SEQ ID NO: 90	334 aa	MW at 36926.0kD
NOV25b, CG59311-02 Protein Sequence	MAATLILEPAGRCCWDEPLRIAVRGLAPEQPVTLRTSLRDEEGALFRAHARYRADASN PGTLGGQGRGPFPGIIDLFGSSRGLCEYRASLLAGHGFAVLALAYFRFEDLPEDLNDV HLEYFEEAVDFMLQHPKVKGPSIALLGFSKGGDLCLSMASFLKGITATVLINACVANT VAPLHYKDMIIPKLVDDLGKVKITKSGFLTFMDTWSNPLEEHNHQSLVPLEKAQVPFL FIVGMDDQSWKSEFYAQIASERLQAHGKERPQIICYPETGHCIDPPYFPPSRASVHAV LGEAIFYGGEPKAHSKAQVDAWQQIQTFFHKHLNGKKSVKHSKI		
	SEQ ID NO: 91	1021 bp	
NOV25c, CG59311-03 DNA Sequence	AGATTGGGATGGCAGCGACGCTGATCCTGGAGCCCGCGGGCCGCTGCTGCTGGGACGA GCCGCTGCGCATCGCAGTGCGCGGCCTGGCCCCGGAGCAGCCAGTCACGCTGCGCACG TCCCTGCGCGACGAAGAGGGCGCGCTCTTCCGGGCCCACGCGCGCTACCGTGCCGACG CCTCTAATCCCGGCACTTTGGGGAGGCCAAGGCAGGGGGCCCTTTCCTGGGATCATTGA TCTGTTTGGGAGCAGCAGGGGCCTTTGTGAATACAGGGCCAGCCTCCTGGCCGGACAT GGTTTTGCTGTGCTTGCCCTGGCTTATTTCAGATTTGAAGACCTCCCCGAAGATCTGA ATGATGTACATCTGGAGTACTTTGAAGAAGCCGTGGACTTTATGCTGCAGCATCCAAA GGTGAAAGGTCCTAGTATTGCGCTTCTTGGATTTTCCAAAGGAGGTGACCTGTGTCTC TCAATGGCTTCTTTCTTGAAGGGCATCACAGCCACTGTACTTATCAATGCCTGTGTAG CCAACACAGTAGCTCCTCTACATTACAAGGATATGATTATTCCTAAACTTGTCGATGA TCTAGGAAAAGTAAAAATCACTAAGTCAGGATTTCTCACTTTTATGGACACTTGGAGC AATCCACTGGAGGAACACAATCACCAAAGTCTTGTTCCATTGGAAAAGGCGCAGGTGC CCTTCTTGTTTATTGTTGGCATGGATGATCAAAGCTGGAAGAGTGAATTCTATGCTCA GATAGCCTCTGAAAGGCTACAAGCTCATGGGAAAGAAAGACCCCAGATAATCTGTTAC CCAGAAACTGGTCACTGTATTGACCCACCTTATTTTCCTCCTTCTAGAGCTTCTGTGC ACGCTGTTTTGGGTGAGGCAATATTCTATGGAGGTGAGCCAAAGGCTCACTCAAAGGC ACAGGTAGATGCCTGGCAGCAAATTCAAACTTTCTTCCATAAACATCTCAATGGTAAA AAATCTGTCAAGCACAGCAAAATATAACATTGTAG		
	ORF Start: ATG at 9	ORF Stop: TAA at 1011	
	SEQ ID NO: 92	334 aa	MW at 36926.0kD
NOV25c, CG59311-03 Protein Sequence	MAATLILEPAGRCCWDEPLRIAVRGLAPEQPVTLRTSLRDEEGALFRAHARYRADASN PGTLGGQGRGPFPGIIDLFGSSRGLCEYRASLLAGHGFAVLALAYFRFEDLPEDLNDV HLEYFEEAVDFMLQHPKVKGPSIALLGFSKGGDLCLSMASFLKGITATVLINACVANT VAPLHYKDMIIPKLVDDLGKVKITKSGFLTFMDTWSNPLEEHNHQSLVPLEKAQVPFL FIVGMDDQSWKSEFYAQIASERLQAHGKERPQIICYPETGHCIDPPYFPPSRASVHAV LGEAIFYGGEPKAHSKAQVDAWQQIQTFFHKHLNGKKSVKHSKI		

178

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 25B.

Table 25B. Comparison of NOV25a against NOV25b through NOV25c.

Protein Sequence	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV25b	154..421 67..334	268/268 (100%) 268/268 (100%)
NOV25c	154..421 67..334	268/268 (100%) 268/268 (100%)

Further analysis of the NOV25a protein yielded the following properties shown in Table 25C.

Table 25C. Protein Sequence Properties NOV25a

PSort analysis:	0.4500 probability located in cytoplasm; 0.3630 probability located in microbody (peroxisome); 0.1958 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

5   A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 25D.

Table 25D. Geneseq Results for NOV25a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41490	Human polypeptide SEQ ID NO 6421 - Homo sapiens, 494 aa. [WO200153312-A1, 26-JUL-2001]	1..421 74..494	288/421 (68%) 347/421 (82%)	e-175
AAM39704	Human polypeptide SEQ ID NO 2849 - Homo sapiens, 483 aa. [WO200153312-A1, 26-JUL-2001]	1..421 63..483	288/421 (68%) 346/421 (81%)	e-175

AAY71112	Human Hydrolase protein-10 (HYDRL-10) - Homo sapiens, 483 aa. [WO200028045-A2, 18-MAY-2000]	1..421 63..483	288/421 (68%) 346/421 (81%)	e-175
AAB93479	Human protein sequence SEQ ID NO:12766 - Homo sapiens, 483 aa. [EP1074617-A2, 07-FEB-2001]	1..421 63..483	287/421 (68%) 346/421 (82%)	e-175
AAY07932	Human secreted protein fragment encoded from gene 81 - Homo sapiens, 182 aa. [WO9918208-A1, 15-APR-1999]	241..421 1..181	181/181 (100%) 181/181 (100%)	e-105

In a BLAST search of public sequence databases, the NOV25a protein was found to have homology to the proteins shown in the BLASTP data in Table 25E.

Table 25E. Public BLASTP Results for NOV25a				
Protein Accession Number	Protein/Organism/Length	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P49753	Peroxisomal acyl-coenzyme A thioester hydrolase 2 (EC 3.1.2.2) (Peroxisomal long-chain acyl-coA thioesterase 2) (ZAP128) - Homo sapiens (Human), 421 aa.	1..421 1..421	288/421 (68%) 347/421 (82%)	e-175
Q9QYR7	Peroxisomal acyl-coenzyme A thioester hydrolase 2 (EC 3.1.2.2) (Peroxisomal long-chain acyl-coA thioesterase 2) (PTE-Ia) - Mus musculus (Mouse), 432 aa.	1..421 12..432	264/421 (62%) 331/421 (77%)	e-157
O88267	Cytosolic acyl coenzyme A thioester hydrolase, inducible (EC 3.1.2.2) (Long chain acyl-CoA thioester hydrolase) (Long chain acyl-CoA hydrolase) (CTE-I) (LACH2) (ACH2) - Rattus norvegicus (Rat), 419 aa.	1..421 1..419	268/421 (63%) 318/421 (74%)	e-153
Q9QYR9	Acyl coenzyme A thioester hydrolase, mitochondrial precursor (EC 3.1.2.2) (Very-long-chain acyl-CoA thioesterase) (MTE-I) - Mus musculus (Mouse), 453 aa.	3..413 44..452	264/411 (64%) 321/411 (77%)	e-153
O55137	Cytosolic acyl coenzyme A thioester	1..413 1..411	262/413 (63%) 319/413 (76%)	e-153

180

	(Long chain acyl-CoA thioester hydrolase) (Long chain acyl-CoA hydrolase) (CTE-I) - Mus musculus (Mouse), 419 aa.			

PFam analysis predicts that the NOV25a protein contains the domains shown in the Table 25F.

<table>
<tr><th colspan="4">Table 25F. Domain Analysis of NOV25a</th></tr>
<tr><th>Pfam Domain</th><th>NOV25a Match Region</th><th>Identities/ Similarities for the Matched Region</th><th>Expect Value</th></tr>
<tr><td colspan="4" align="center">No Significant Matches Found</td></tr>
</table>

Example 26.

The NOV26 clone was analyzed, and the nucleotide and predicted polypeptide
sequences are shown in Table 26A.

<table>
<tr><th colspan="3">Table 26A. NOV26 Sequence Analysis</th></tr>
<tr><td></td><td>SEQ ID NO: 93</td><td>1375 bp</td></tr>
<tr><td>NOV26a,
CG59309-01 DNA Sequence</td><td colspan="2">GGGACGCCGGACGCCGTCCGGACATTCGGCGCGCTTGCCACGATCTTGGACGGGTCTC
GGGCCTCGACCTTTGAATTCCCCGCTCCGGCTCCAAGATGTCAGCAACGCTGATCCTG
GAGCCCCCCAGGCCGCTGCTGCTGGAACGAGCCGGTGCGCATTGCCGTGCGCGGCCTGG
CCCCGGAGCAGCGGGTTACGCTGCGCGCGCGTCCCTGCGCGACGAGAAGGGCGCGCTCTT
CCGGGCCCACGCGCGCTACTGCGCCGACGCCCGCGGCGAGCTGGACCTGGAGCGCGCA
CCCGCGCTGGGCGGCAGCTTCGCGGGACTCGAGCCCATGGGGCTGCTCTGGGCCCTGG
AACCCGAGAAGCCTTTTGGCGCGTTCCTGAAGCGGGACGTACAGATTCCTTTTGTCGT
GGAGTTGGAGGTGCTGGACGGCCACGACCCCGAGCCTGGACGGCTGCTGTGCCAGGCG
CAGCACGAGCGCCACTTCCTCCCGCCAGGGGTGCGGCGCCAGTCGGTGCGAGCGGGCC
GGGTGCGCGCCACGCTCTTCCTGCCGCCAGGTGAGCCTGGACCCTTCCCAGGGATCAT
TGACATCTTTGGTATTGGAGGGGGCCTCTTGGAATATCGAGCCAGCCTCCTTGCTGGC
CATGGCTTTGCCACGTTGGCTCTAGCTTATTATAACTTTGAAGATCTCCCCAATAACA
TGGACAACATATCCCTGGAGTACTTCGAAGAAGCCGTATGCTACATGCTTCAACATCC
CCAGGTTAAAGGCCCAGGCATTGGGCTTTTGGGCATTTCTCTAGGAGCTGATATTTGT
CTCTCAATGGCCTCATTCTTGAAGAATGTCTCAGCCACAGTTTCCATCAATGGATCTG
GGATCAGTGGGAACACAGCCATCAACTATAAGCACAGTAGCATTCCACCATTGGGCTA
TGACCTGAGGAGAATCAAGGTAGCTTTCTCAGGCCTCGTGGACATTGTGGATATAAGG
AATGCTCTCGTAGGAGGGTACAAGAACCCCAGCATGATTCCAATAGAGAAGGCCCAGG
GGCCCATCCTGCTCATTGTTGGTCAGGATGACCATAACTGGAGAAGTGAGTTGTATGC
CCAAACAGTCTCTGAACGGTTACAGGCCCATGGAAAGGAAAAACCCCAGATCATCTGT
TACCCTGGGACTGGGCATTACATCGAGCCTCCTTACTTCCCCCTGTGCCCAGCTTCCC
TTCACAGATTACTGAACAAACATGTTATATGGGGTGGGGAGCCCAGGGCTCATTCTAA
GGCCCAGGAAGATGCCTGGAAGCAAATTCTAGCCTTCTTCTGCAAACACCTGGGAGGT
ACCCAGAAAACAGCTGTCCCTAAATTGTAATGCATTTGTCT</td></tr>
<tr><td></td><td>ORF Start: ATG at 96</td><td>ORF Stop: TAA at 1362</td></tr>
<tr><td></td><td>SEQ ID NO: 94</td><td>422 aa | MW at 46455.1kD</td></tr>
<tr><td>NOV26a,
CG59309-01 Protein Sequence</td><td colspan="2">MSATLILEPPGRCCWNEPVRIAVRGLAPEQRVTLRASLRDEKGALFRAHARYCADARG
ELDLERAPALGGSFAGLEPMGLLWALEPEKPFWRFLKRDVQIPFVVELEVLDGHDPEP
GRLLCQAQHERHFLPPGVRRQSVRAGRVRATLFLPPGEPGPFPGIIDIFGIGGGLLEY
RASLLAGHGFATLALAYYNFEDLPNNMDNISLEYFEEAVCYMLQHPQVKGPGIGLLGI
SLGADICLSMASFLKNVSATVSINGSGISGNTAINYKHSSIPPLGYDLRRIKVAFSGL
VDIVDIRNALVGGYKNPSMIPIEKAQGPILLIVGQDDHNWRSELYAQTVSERLQAHGK
EKPQIICYPGTGHYIEPPYFPLCPASLHRLLNKHVIWGGEPRAHSKAQEDAWKQILAF
FCKHLGGTQKTAVPKL</td></tr>
</table>

Further analysis of the NOV26a protein yielded the following properties shown in Table 26B.

Table 26B. Protein Sequence Properties NOV26a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.2585 probability located in lysosome (lumen); 0.1940 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV26a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 26C.

Table 26C. Geneseq Results for NOV26a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41490	Human polypeptide SEQ ID NO 6421 - Homo sapiens, 494 aa. [WO200153312-A1, 26-JUL-2001]	1..422 74..494	296/422 (70%) 341/422 (80%)	e-179
AAM39704	Human polypeptide SEQ ID NO 2849 - Homo sapiens, 483 aa. [WO200153312-A1, 26-JUL-2001]	1..422 63..483	296/422 (70%) 341/422 (80%)	e-179
AAY71112	Human Hydrolase protein-10 (HYDRL-10) - Homo sapiens, 483 aa. [WO200028045-A2, 18-MAY-2000]	1..422 63..483	296/422 (70%) 341/422 (80%)	e-179
AAB93479	Human protein sequence SEQ ID NO:12766 - Homo sapiens, 483 aa. [EP1074617-A2, 07-FEB-2001]	1..422 63..483	295/422 (69%) 340/422 (79%)	e-178
AAY07932	Human secreted protein fragment encoded from gene 81 - Homo sapiens, 182 aa. [WO9918208-A1, 15-APR-1999]	242..422 1..181	93/181 (51%) 123/181 (67%)	2e-48

In a BLAST search of public sequence databases, the NOV26a protein was found to have homology to the proteins shown in the BLASTP data in Table 26D.

Table 26D. Public BLASTP Results for NOV26a				
Protein Accession Number	Protein/Organism/Length	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9QYR8	PEROXISOMAL LONG CHAIN ACYL-COA THIOESTERASE IB - Mus musculus (Mouse), 421 aa.	1..422 1..421	312/422 (73%) 362/422 (84%)	0.0
P49753	Peroxisomal acyl-coenzyme A thioester hydrolase 2 (EC 3.1.2.2) (Peroxisomal long-chain acyl-coA thioesterase 2) (ZAP128) - Homo sapiens (Human), 421 aa.	1..422 1..421	296/422 (70%) 341/422 (80%)	e-178
Q9QYR7	Peroxisomal acyl-coenzyme A thioester hydrolase 2 (EC 3.1.2.2) (Peroxisomal long-chain acyl-coA thioesterase 2) (PTE-Ia) - Mus musculus (Mouse), 432 aa.	1..422 12..432	281/424 (66%) 333/424 (78%)	e-163
O55137	Cytosolic acyl coenzyme A thioester hydrolase, inducible (EC 3.1.2.2) (Long chain acyl-CoA thioester hydrolase) (Long chain acyl-CoA hydrolase) (CTE-I) - Mus musculus (Mouse), 419 aa.	1..422 1..419	275/423 (65%) 330/423 (78%)	e-162
O88267	Cytosolic acyl coenzyme A thioester hydrolase, inducible (EC 3.1.2.2) (Long chain acyl-CoA thioester hydrolase) (Long chain acyl-CoA hydrolase) (CTE-I) (LACH2) (ACH2) - Rattus norvegicus (Rat), 419 aa.	1..422 1..419	276/423 (65%) 329/423 (77%)	e-162

PFam analysis predicts that the NOV26a protein contains the domains shown in the Table 26E.

Table 26E. Domain Analysis of NOV26a			
Pfam Domain	NOV26a Match Region	Identities/ Similarities for the Matched Region	Expect Value
DLH: domain 1 of 2	144..188	17/52 (33%) 32/52 (62%)	63
DLH: domain 2 of 2	394..411	9/18 (50%) 13/18 (72%)	2.6

Example 27.

The NOV27 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 27A.

Table 27A. NOV27 Sequence Analysis		
	SEQ ID NO: 95	1333 bp
NOV27a, CG57364-01 DNA Sequence	CCTGGCCCCCAAGCTCCCCACTCTGGTGCCCCGAGCAGCCCTGTGGGCAAGCAGCCGC CGCCATGGCCGAGCACCTGGAGCTGCTGGCAGAGATGCCCATGGTGGGCAGGATGAGC ACACAGGAGCGGCTGAAGCATGCCCAGAAGCGGCGCGCCCAGCAGGTGAAGATGTGGG CCCAGGCTGAGAAGGAGGCCCAGGGCAAGAAGGGTCCTGGGGAGCGTCCCCGGAAGGA GGCAGCCAGCCAAGGGCTCCTGAAGCAGGTCCTCTTCCCTCCCAGTGTTGTCCTTCTG GAGGCCGCTGCCCGAAATGACCTGGAAGAAGTCCGCCAGTTCCTTGGGAGTGGGGTCA GCCCTGACTTGGCCAACGAGGACGGCCTGACGGCCCTGCACCAGTGCTGCATTGATGA TTTCCGAGAGATGGTGCAGCAGCTCCTGGAGGCTGGGGCCAACATCAATGCCTGTGAC AGTGAGTGCTGGACGCCTCTGCATGCTGCGGCCACCTGCGGCCACCTGCACCTGGTGG AGCTGCTCATCGCCAGTGGCGCCAATCTCCTGGCGGTCAACACCGACGGGAACATGCC CTATGACCTGTGTGATGATGAGCAGACGCTGGACTGCCTGGAGACTGCCATGGCCGAC CGTGGCATCACCCAGGACAGCATCGAGGCCGCCCGGGCCGTGCCAGAACTGCGCATGC TGGACGACATCCGGAGCCGGCTGCAGGCCGGGGCAGACCTCCATGCCCCCCTGGACCA CGGGGCCACGCTGCTGCACGTCGCAGCCGCCAACGGGTTCAGCGAGGCGGCTGCCCTG CTGCTGGAACACCGAGCCAGCCTGAGCGCTAAGGACCAAGACGGCTGGGAGCCGCTGC ACGCCGCGGCCTACTGGGGCCAGGTGCCCCTGGTGGAGCTGCTCGTGGCCGCACGGGGC CGACCTGAACGCAAAGTCCCTGATGGACGAGACGCCCCTTGATGTGTGCGGGGACGAG GAGGTGCGGGCCAAGCTGCTGGAGCTGAAGCACAAGCACGACGCCCTCCTGCGCGCCC AGAGCCGCCAGCGCTCCTTGCTGCGCCGCCGCACCTCCAGCGCCGGCAGCCGCGGGAA GGTGGTGAGGCGGGATGAGCCTAACCCAGCGCAGCGGCTGACGCATGTCCCAGAAGCG GCGCGCCCAGCAGGTGAAGATGTGGGCCCAGGCTGAGAAGGAGGCCCAGGGCAAGAAG GGTCCTGGGGAGCGTCCCCGGAAGGAGGCAGCCAGCCAAGGGCTCCTGAAGCAGGTCC TCTTCCCTCCCAGTGTTGTCCTTCTGGAGGCCGCTGCCCGAAATGACCTGGAAGAAG
	ORF Start: ATG at 63	ORF Stop: TGA at 1194
	SEQ ID NO: 96	377 aa	MW at 41019.9kD
NOV27a, CG57364-01 Protein Sequence	MAEHLELLAEMPMVGRMSTQERLKHAQKRRAQQVKMWAQAEKEAQGKKGPGERPRKEA ASQGLLKQVLFPPSVVLLEAAARNDLEEVRQFLGSGVSPDLANEDGLTALHQCCIDDF REMVQQLLEAGANINACDSECWTPLHAAATCGHLHLVELLIASGANLLAVNTDGNMPY DLCDDEQTLDCLETAMADRGITQDSIEAARAVPELRMLDDIRSRLQAGADLHAPLDHG ATLLHVAAANGFSEAAALLLEHRASLSAKDQDGWEPLHAAAYWGQVPLVELLVAHGAD LNAKSLMDETPLDVCGDEEVRAKLLELKHKHDALLRAQSRQRSLLRRRTSSAGSRGKV VRRDEPNPAQRLTHVPEAARPAGEDVGPG	

Further analysis of the NOV27a protein yielded the following properties shown in Table 27B.

Table 27B. Protein Sequence Properties NOV27a	
PSort analysis:	0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1547 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 27C.
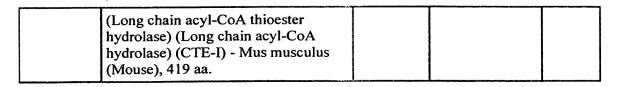
184

Table 27C. Geneseq Results for NOV27a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM40636	Human polypeptide SEQ ID NO 5567 - Homo sapiens, 440 aa. [WO200153312-A1, 26-JUL-2001]	89..351 1..263	262/263 (99%) 263/263 (99%)	e-151
AAM38850	Human polypeptide SEQ ID NO 1995 - Homo sapiens, 410 aa. [WO200153312-A1, 26-JUL-2001]	119..351 1..233	233/233 (100%) 233/233 (100%)	e-132
AAM78864	Human protein SEQ ID NO 1526 - Homo sapiens, 567 aa. [WO200157190-A2, 09-AUG-2001]	1..351 1..348	209/351 (59%) 265/351 (74%)	e-118
ABB11817	Human KIAA0823 protein homologue, SEQ ID NO:2187 - Homo sapiens, 536 aa. [WO200157188-A2, 09-AUG-2001]	45..354 3..318	173/316 (54%) 226/316 (70%)	3e-94
AAM79848	Human protein SEQ ID NO 3494 - Homo sapiens, 536 aa. [WO200157190-A2, 09-AUG-2001]	45..354 3..318	173/316 (54%) 226/316 (70%)	3e-94

In a BLAST search of public sequence databases, the NOV27a protein was found to have homology to the proteins shown in the BLASTP data in Table 27D.

Table 27D. Public BLASTP Results for NOV27a				
Protein Accession Number	Protein/Organism/Length	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96I34	UNKNOWN (PROTEIN FOR MGC:14333) - Homo sapiens (Human), 528 aa.	1..351 1..351	351/351 (100%) 351/351 (100%)	0.0
Q923M0	MYOSIN PHOSPHATASE TARGETING SUBUNIT 3 MYPT3 - Mus musculus (Mouse), 524 aa (fragment).	1..351 1..351	301/351 (85%) 320/351 (90%)	e-171

185

AAL62093	PROTEIN PHOSPHATASE 1 REGULATORY SUBUNIT 16B - Mus musculus (Mouse), 568 aa.	1..351 1..348	210/351 (59%) 266/351 (74%)	e-118
Q95N27	CAAX BOX PROTEIN TIMAP - Bos taurus (Bovine), 568 aa.	1..351 1..348	210/351 (59%) 266/351 (74%)	e-118
Q96T49	CAAX BOX PROTEIN TIMAP - Homo sapiens (Human), 567 aa.	1..351 1..348	209/351 (59%) 265/351 (74%)	e-117

PFam analysis predicts that the NOV27a protein contains the domains shown in the Table 27E.

<table>
<tr><th colspan="4">Table 27E. Domain Analysis of NOV27a</th></tr>
<tr><th>Pfam Domain</th><th>NOV27a Match Region</th><th>Identities/ Similarities for the Matched Region</th><th>Expect Value</th></tr>
<tr><td>ank: domain 1 of 5</td><td>70..102</td><td>8/33 (24%) 20/33 (61%)</td><td>99</td></tr>
<tr><td>ank: domain 2 of 5</td><td>103..135</td><td>16/33 (48%) 26/33 (79%)</td><td>7.1e-08</td></tr>
<tr><td>ank: domain 3 of 5</td><td>136..168</td><td>15/33 (45%) 26/33 (79%)</td><td>2.9e-07</td></tr>
<tr><td>ank: domain 4 of 5</td><td>231..263</td><td>16/33 (48%) 24/33 (73%)</td><td>2e-06</td></tr>
<tr><td>ank: domain 5 of 5</td><td>264..296</td><td>16/33 (48%) 27/33 (82%)</td><td>2.7e-08</td></tr>
</table>

Example 28.

The NOV28 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 28A.

<table>
<tr><th colspan="3">Table 28A. NOV28 Sequence Analysis</th></tr>
<tr><td></td><td>SEQ ID NO: 97</td><td>1719 bp</td></tr>
<tr><td>NOV28a, CG59348-01 DNA Sequence</td><td colspan="2">CGGGCACAGGCTCACCCTCGAGTGGCACAGGAATCCCAGGTAGATGACGGCGGCCGCG
GCTGGTGCTGCAGGGTCGGCAGCTCCCGCGGCAGCGGCCGGCGCCCCGGGATCTGGGG
GCGCACCCTCAGGGTCGCAGGGGGTGCTGATCGGGGACAGGCTGTACTCCGGGGTGCT
CATCACCTTGGAGAACTGCCTCCTGCCTGACGACAAGCTCCGTTTCACGCCGTCCATG
TCGAGCGGCCTCGACACCGACACAGAGACCGACCTCCGCGTGGTGGGCTGCGAGCTCA
TCCAGGCGGCCGGTATCCTGCTCCGCCTGCCGCAGGTGGCCATGGCTACCGGGCAGGT
GTTGTTCCAGCGGTTCTTTTATACCAAGTCCTTCGTGAAGCACTCCATGGAGCATGTG
TCAATGGCCTGTGTCCACCTGGCTTCCAAGATAGAAGAGGCCCCAAGACGCATACGGG
ACGTCATCAATGTGTTTCACCGCCTTCGACAGCTGAGAGACAAAAAGAAGCCCGTGCC
TCTACTACTGGATCAAGATTATGTTAATTTAAAGAACCAAATTATAAAGGCGGAAAGA
CGAGTTCTCAAAGAGTTGGGTTTCTGCGTCCATGTGAAGCATCCTCATAAGATAATCG
TTATGTACCTTCAGGTGTTAGAGTGTGAGCGTAACCAACACCTGGTCCAGACCTCATG
GAATTACATGAACGACAGCCTTCGCACCGACGTCTTCGTGCGGTTCCAGCCAGAGAGC
ATCGCCTGTGCCTGCATTTATCTTGCTGCCCGGACGCTGGAGATCCCTTTGCCCAATC
GTCCCCATTGGTTTCTTTTGTTTGGAGCAACTGAAGAAGAAATTCAGGAAATCTGCTT</td></tr>
</table>

| | AAAGATCTTGCAGCTTTATGCTCGGAAAAAGGTTGATCTCACACACCTGGAGGGTGAA
GTGGAAAAAAGAAAGCACGCTATCGAAGAGGCAAAGGCCAAGCCCGGGGCCTGTTGC
CTGGGGGCACACAGGTGCTGGATGGTACCTCGGGGTTCTCTCCTGCCCCCAAGCTGGT
GGAATCCCCCAAAGAAGGTAAAGGGAGCAAGCCTTCCCCACTGTCTGTGAAGAACACC
AAGAGGAGGCTGGAGGGCGCCAAGAAAGCCAAGGCGGACAGCCCCGTGAACGGCTTGC
CAAAGGGGCGAGAGAGTCGGAGTCGGAGCCGGAGCCGTGAGCAGAGCTACTCGAGGTC
CCCATCCCGATCAGCGTCTCCTAAGAGGAGGAAAAGTGACAGCGGCTCCACATCTGGT
GGGTCCAAGTCGCAGAGCCGCTCCCGGAGCAGGAGTGACTCCCCACCGAGACAGGCCC
CCCGCAGCGCTCCCTACAAAGGCTCTGAGATTCGGGGCTCCCGGAAGTCCAAGGACTG
CAAGTACCCCCAGAAGCCACACAAGTCTCGGAGCCGGAGTTCTTCCCGTTCTCGAAGC
AGGTCACGGGAGCGGGCGGATAATCCGGGAAAATACAAGAAGAAAAGTCATTACTACA
GAGATCAGCGACGAGAGCGCTCGAGGTCGTATGAACGCACAGGCCGTCGCTATGAGCG
GGACCACCCTGGGCACAGCAGGCATCGGAGGTGACACGTGCTTCAGACCGGTCTGGGG
TGCGGCGCACACCTGGGCCCGTGCAGGGCTCAGCTCGGCAGCAGCTCTGAGGGCAGCT
CAATGAAAAAGTGAATGCACACGCCCTTGTTGGCGTG |

ORF Start: ATG at 44	ORF Stop: TGA at 1598	

SEQ ID NO: 98	518 aa	MW at 58034.5kD

NOV28a, CG59348-01 Protein Sequence	MTAAAAGAAGSAAPAAAAGAPGSGGAPSGSQGVLIGDRLYSGVLITLENCLLPDDKLR FTPSMSSGLDTDTETDLRVVGCELIQAAGILLRLPQVAMATGQVLFQRFFYTKSFVKH SMEHVSMACVHLASKIEEAPRRIRDVINVFHRLRQLRDKKKPVPLLLDQDYVNLKNQI IKAERRVLKELGFCVHVKHPHKIIVMYLQVLECERNQHLVQTSWNYMNDSLRTDVFVR FQPESIACACIYLAARTLEIPLPNRPHWFLLFGATEEEIQEICLKILQLYARKKVDLT HLEGEVEKRKHAIEEAKAQARGLLPGGTQVLDGTSGFSPAPKLVESPKEGKGSKPSPL SVKNTKRRLEGAKKAKADSPVNGLPKGRESRSRSRSREQSYSRSPSRSASPKRRKSDS GSTSGGSKSQSRSRSRSDSPPRQAPRSAPYKGSEIRGSRKSKDCKYPQKPHKSRSRSS SRSRSRSRERADNPGKYKKKSHYYRDQRRERSRSYERTGRRYERDHPGHSRHRR

Further analysis of the NOV28a protein yielded the following properties shown in Table 28B.

Table 28B. Protein Sequence Properties NOV28a	
Psort analysis:	0.5500 probability located in endoplasmic reticulum (membrane); 0.2400 probability located in nucleus; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV28a protein against the Geneseq database, a proprietary
database that contains sequences published in patents and patent publication, yielded
several homologous proteins shown in Table 28C.

Table 28C. Geneseq Results for NOV28a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM94028	Human stomach cancer expressed polypeptide SEQ ID NO 126 - Homo sapiens, 298 aa. [WO200109317-A1, 08-FEB-2001]	221..518 1..298	298/298 (100%) 298/298 (100%)	e-172

AAG64403	Human paneth cell enhanced expression-like protein - Homo sapiens, 298 aa. [WO200138372-A1, 31-MAY-2001]	221..518 1..298	298/298 (100%) 298/298 (100%)	e-172
AAB94641	Human protein sequence SEQ ID NO:15526 - Homo sapiens, 298 aa. [EP1074617-A2, 07-FEB-2001]	221..518 1..298	298/298 (100%) 298/298 (100%)	e-172
AAM78533	Human protein SEQ ID NO 1195 - Homo sapiens, 526 aa. [WO200157190-A2, 09-AUG-2001]	2..518 8..526	316/526 (60%) 390/526 (74%)	e-168
AAB94371	Human protein sequence SEQ ID NO:14909 - Homo sapiens, 526 aa. [EP1074617-A2, 07-FEB-2001]	2..518 8..526	316/526 (60%) 390/526 (74%)	e-168

In a BLAST search of public sequence databases, the NOV28a protein was found to have homology to the proteins shown in the BLASTP data in Table 28D.

Table 28D. Public BLASTP Results for NOV28a				
Protein Accession Number	Protein/Organism/Length	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96S94	HYPOTHETICAL 58.1 KDA PROTEIN - Homo sapiens (Human), 520 aa.	3..518 5..520	516/516 (100%) 516/516 (100%)	0.0
Q9JJA7	BRAIN CDNA, CLONE MNCB-5160, SIMILAR TO MUS MUSCULUS PANETH CELL ENHANCED EXPRESSION PCEE MRNA - Mus musculus (Mouse), 518 aa.	1..518 1..518	466/519 (89%) 482/519 (92%)	0.0
Q9UK58	CYCLIN L ANIA-6A - Homo sapiens (Human), 526 aa.	2..518 8..526	316/526 (60%) 390/526 (74%)	e-167
Q9R1Q2	CYCLIN ANIA-6A - Rattus norvegicus (Rat), 527 aa.	2..518 9..527	312/526 (59%) 391/526 (74%)	e-165
Q9WV44	CYCLIN ANIA-6A - Mus musculus (Mouse), 531 aa.	3..518 15..531	314/526 (59%) 385/526 (72%)	e-162

PFam analysis predicts that the NOV28a protein contains the domains shown in the Table 28E.

188

Table 28E. Domain Analysis of NOV28a

Pfam Domain	NOV28a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cyclin: domain 1 of 1	46..190	28/163 (17%) 86/163 (53%)	0.0022
Srg: domain 1 of 1	221..230	4/10 (40%) 10/10 (100%)	6.7
transcript_fac2: domain 1 of 1	235..253	12/19 (63%) 15/19 (79%)	0.86
cyclin_C: domain 1 of 1	196..311	22/139 (16%) 65/139 (47%)	2.6

Example 29.

The NOV29 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 29A.

Table 29A. NOV29 Sequence Analysis

	SEQ ID NO: 99	1069 bp	
NOV29a, CG59245-01 DNA Sequence	CGGGGCCTGGTCGGCAGCTGGGCCGCC**ATG**GAGTCCACGCTGGGCGCGGGCATCGTGA TAGCCGAGGCGCTACAGAACCAGCTAGCCTGGCTGGAGAACGTGTGGCTCTGGATCAC CTTTCTGGGCGATCCCAAGATCCTCTTTCTGTTCTACTTCCCCGCGGCCTACTACGCC TCCCGCCGTGTGGGCATCGCGGTGCTCTGGATCAGCCTCATCACCGAGTGGCTCAACC TCATCTTCAAGTGGTTTCTTTTTGGAGACAGGCCCTTTTGGTGGGTCCATGAGTCTGG TTACTACAGCCAGGCTCCAGCCCAGGTTCACCAGTTCCCCTCTTCTTGTGAGACTGGT CCAGGTGGCAGCCCTTCTGGACACTGCATGATCACAGGAGCAGCCCTCTGGCCCATAA TGACGGCCCTGTCTTCGCAGGTGCGCTGGGTAAGGGTGATGCCTAGCCTGGCTTATTG CACCTTCCTTTTGGCGGTTGGCTTGTCGCGAATCTTCATCTTAGCACATTTCCCTCAC CAGGTGCTGGCTGGCCTAATAACTGGTTGGCTGATGACTCCCCGAGTGCCTATGGAGC GGGAGCTAAGCTTCTATGGGTTGACTGCACTGGCCCTCATGCTAGGCACCAGCCTCAT CTATTGGACCCTCTTTACACTGGGCCTGGATCTTTCTTGGTCCATCAGCCTAGCCTTC AAGTGGTGTGAGCGGCCTGAGTGGATACACGTGGATAGCCGGCCCTTTGCCTCCCTGA GCCGTGACTCAGGGGCTGCCCTGGGCCTGGGCATTGCCTTGCACTCTCCCTGCTATGC CCAGGTGCGTCGGGCACAGCTGGGGAAATGGCCAGAAGATAGCCTGCCTTGTGCTGGCC ATGGGGCTGCTGGGCCCCCTGGACTGGCTGGGCCACCCCCCTCAGATCAGCCTCTTCT ACATTTTCAATTTCCTCAAGTACACCCTCTGGCCATGCCTAGTCCTGGCCCTCGTGCC CTGGGCAGTGCACATGTTCAGTGCCCAGGAAGCACCGCCCATCCACTCTTCC**TGA**_CTT CTTGTGTGCCTCCCTTTCCTTTCCC_		
	ORF Start: ATG at 28	ORF Stop: TGA at 1039	
	SEQ ID NO: 100	337 aa	MW at 37808.0kD
NOV29a, CG59245-01 Protein Sequence	MESTLGAGIVIAEALQNQLAWLENVWLWITFLGDPKILFLFYFPAAYYASRRVGIAVL WISLITEWLNLIFKWFLFGDRPFWWVHESGYYSQAPAQVHQFPSSCETGPGGSPSGHC MITGAALWPIMTALSSQVRWVRVMPSLAYCTFLLAVGLSRIFILAHFPHQVLAGLITG WLMTPRVPMERELSFYGLTALALMLGTSLIYWTLFTLGLDLSWSISLAFKWCERPEWI HVDSRPFASLSRDSGAALGLGIALHSPCYAQVRRAQLGNGQKIACLVLAMGLLGPLDW LGHPPQISLFYIFNFLKYTLWPCLVLALVPWAVHMFSAQEAPPIHSS		
	SEQ ID NO: 101	1386 bp	
NOV29b, CG59245-02 DNA Sequence	TGAGTCTGTACTTTCCGCCCTGGAGCAAGCCGGGGCCTGGTCGGCAGCTGGGCCGCCA TGGAGTCCACGCTGGGCGCGGGCATCGTGATAGCCGAGGCGCTACAGAACCAGCTAGC CTGGCTGGAGAACGTGTGGCTCTGGATCACCTTTCTGGGCGATCCCAAGATCCTCTTT CTGTTCTACTTCCCCGCGGCCTACTACGCCTCCCGCCGTGTGGGCATCGCGGTGCTCT GGATCAGCCTCATCACCGAGTGGCTCAACCTCATCTTCAAGTGGTTTCTTTTTGGAGA CAGGCCCTTTTGGTGGGTCCATGAGTCTGGTTACTACAGCCAGGCTCCAGCCCAGGTT CACCAGTTCCCCTCTTCTTGTGAGACTGGTCCAGGCAGCCCTTCTGGACACTGCATGA		

189

	TCACAGGAGCAGCCCTCTGGCCCATAATGACGGCCCTGTCTTCGCAGGTGGCCACTCG GGCCCGCAGCCGCTGGGTAAGGGTGATGCCTAGCCTGGCTTATTGCACCTTCCTTTTG GCCGGTTGGCTTGTCGCGAATCTTCATCTTAGCACATTTCCCTCACCAGGTGCTGGCTG GCCTAATAACTGGCGCTGTCCTGGGCTGGCTGATGACTCCCCGAGTGCCTATGGAGCG GGAGCTAAGCTTCTATGGGTTGACTGCACTGGCCCTCATGCTAGGCACCAGCCTCATC TATTGGACCCTCTTTACACTGGGCCTGGATCTTTCTTGGTCCATCAGCCTAGCCTTCA AGTGGTGTGAGCGGCCTGAGTGGATACACGTGGATAGCCGGCCCTTTGCCTCCCTGAG CCGTGACTCAGGGGCTGCCCTGGGCCTGGGCATTGCCTTGCACTCTCCCTGCTATGCC CAGGTGCGTCGGGCACAGCTGGGAAATGGCCAGAAGATAGCCTGCCTTGTGCTGGCCA TGGGGCTGCTGGGCCCCCTGGACTGGCTGGGCCACCCCCCTCAGATCAGCCTCTTCTA CATTTTCAATTTCCTCAAGTACACCCTCTGGCCATGCCCAGTCCTGGCCCTCGTGCCC TGGGCAGTGCACATGTTCAGTGCCCAGGAAGCACCGCCCATCCACTCTTCCTGACTTC TTGTGTGCCTCCCTTTCCTTTCCCTCCCACAAAGCCAACACTCTGTGACCACCACACT CCAGGAGGCAGCCCCATCCCCTTCCAGCCCCTAAGTAGGCCCTCCCCTCCCTAAATCT GCTTCCGCACCACCTGGTCTTAGCCCCAAAGATGGGCCTTCTCTCTCCCAGATAAGTT GGTCCTCCCTCTGCCTTTCCTCTCAAGCCCCCAAAGAGCAAAGGCAACAGCAAGACCA GCGGGTTCTTGCAACACTGTGAGGGGCAGCCAGGGCGGAAAGTACAGACTCA
	ORF Start: ATG at 58 **ORF Stop: TGA at 1096**
	SEQ ID NO: 102 346 aa MW at 38718.0kD
NOV29b, CG59245-02 Protein Sequence	MESTLGAGIVIAEALQNQLAWLENVWLWITFLGDPKILFLFYFPAAYYASRRVGIAVL WISLITEWLNLIFKWFLFGDRPFWWVHESGYYSQAPAQVHQFPSSCETGPGSPSGHCM ITGAALWPIMTALSSQVATRARSRWVRVMPSLAYCTFLLAVGLSRIFILAHFPHQVLA GLITGAVLGWLMTPRVPMERELSFYGLTALALMLGTSLIYWTLFTLGLDLSWSISLAF KWCERPEWIHVDSRPFASLSRDSGAALGLGIALHSPCYAQVRRAQLGNGQKIACLVLA MGLLGPLDWLGHPPQISLFYIFNFLKYTLWPCPVLALVPWAVHMFSAQEAPPIHSS

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 29B.

Table 29B. Comparison of NOV29a against NOV29b.		
Protein Sequence	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV29b	1..337 1..346	335/347 (96%) 335/347 (96%)

Further analysis of the NOV29a protein yielded the following properties shown in Table 29C.

Table 29C. Protein Sequence Properties NOV29a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 41 and 42

A search of the NOV29a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 29D.

190

Table 29D. Geneseq Results for NOV29a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV29a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAM79500	Human protein SEQ ID NO 3146 - Homo sapiens, 382 aa. [WO200157190-A2, 09-AUG-2001]	1..337 37..382	336/347 (96%) 336/347 (96%)	0.0
AAB42637	Human ORFX ORF2401 polypeptide sequence SEQ ID NO:4802 - Homo sapiens, 377 aa. [WO200058473-A2, 05-OCT-2000]	1..337 31..377	328/348 (94%) 328/348 (94%)	0.0
AAB85355	Human phosphatase (PP) (clone ID 1269556CD1) - Homo sapiens, 385 aa. [WO200153469-A2, 26-JUL-2001]	1..305 1..314	297/315 (94%) 298/315 (94%)	e-174
AAM78516	Human protein SEQ ID NO 1178 - Homo sapiens, 404 aa. [WO200157190-A2, 09-AUG-2001]	1..337 125..404	266/341 (78%) 272/341 (79%)	e-146
AAB25679	Human secreted protein sequence encoded by gene 15 SEQ ID NO:68 - Homo sapiens, 141 aa. [WO200043495-A2, 27-JUL-2000]	198..337 1..140	140/140 (100%) 140/140 (100%)	6e-81

In a BLAST search of public sequence databases, the NOV29a protein was found to have homology to the proteins shown in the BLASTP data in Table 29E.

Table 29E. Public BLASTP Results for NOV29a				
Protein Accession Number	**Protein/Organism/Length**	**NOV29a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAH21574	HYPOTHETICAL 38.7 KDA PROTEIN - Homo sapiens (Human), 346 aa.	1..337 1..346	336/347 (96%) 336/347 (96%)	0.0
Q9BUM1	HYPOTHETICAL 40.1 KDA PROTEIN - Homo sapiens (Human), 360 aa (fragment).	1..337 15..360	336/347 (96%) 336/347 (96%)	0.0

191

O42153	Glucose-6-phosphatase (EC 3.1.3.9) (G6Pase) (G-6-Pase) - Haplochromis nubilus, 352 aa.	8..323 8..339	127/333 (38%) 184/333 (55%)	1e-59
Q98UF8	GLUCOSE-6-PHOSPHATASE - Sparus aurata (Gilthead sea bream), 350 aa.	8..323 8..337	123/333 (36%) 185/333 (54%)	2e-57
Q9Z186	GLUCOSE-6-PHOSPHATASE - Mus musculus (Mouse), 355 aa.	7..325 7..345	128/343 (37%) 188/343 (54%)	5e-56

PFam analysis predicts that the NOV29a protein contains the domains shown in the Table 29F.

Table 29F. Domain Analysis of NOV29a			
Pfam Domain	NOV29a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PAP2: domain 1 of 1	51..190	38/175 (22%) 95/175 (54%)	0.00037

Example 30.

The NOV30 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 30A.

Table 30A. NOV30 Sequence Analysis		
	SEQ ID NO: 103	1624 bp
NOV30a, CG59241-01 DNA Sequence	ATGGAACTGAAGGCCGAGGAGGAGGAGGTGGGTGGCGTCCAGCCGGTGGACTTGGTGG CCTTTGCCAACAGCTGCACCCTCCATGGCACCAACCACATTTTTGTGGAGGGGGGTCC AGGGCCAAGGCAGGTGCTGTGGGCGGTGGCCTTTGTCCTGGCACTGGGTGCCTTCCTG TGCCAGGTAGGGGACCGCGTTGCTTATTACCTCAGCTACCCACACGTGACCCTTCTAA ACGAAGTGGCCACCACGGAGCTGGCCTTCCCGGCAGTCACCCTCTGCAACACTAATGC TGTGCGGCTGTCCCAGCTCAGCTACCCTGACTTGCTTTATTTGGCCCCCATGCTGGGA CTGGATGAAAGTGATGACCCCGGGGTGCCCCTCGCTCCACCGGGCCCTGAGGCCTTCT CTGGGGAGCCCTTTAACCTGCACCGCTTCTACAATCGCTCCTGCCACCGGCTGGAGGA CATGCTGCTCTATTGCTCCTACCAAGGGGGACCCTGCGGCCCTCACAACTTCTCAGTG GTGTTCACACGCTATGGAAAGTGCTACACGTTCAACTCGGGCCGAGATGGGCGGCCGC GGCTGAAGACCATGAAGGGTGGGACGGGCAATGGGCTGGAAATCATGCTGGACATCCA GCAGGACGAGTACCTGCCTGTGTGGGGGGAGACTGACGAGACGTCCTTCGAAGCAGGC ATCAAAGTGCAGATCCATAGTCAGGATGAACCTCCTTTCATCGACCAGCTGGGCTTTG GCGTGGCCCCAGGCTTCCAGACCTTTGTGGCCTGCCAGGAGCAGCGGATCTACCTGCC CCCACCCTGGGGCACCTGCAAAGCTGTTACCATGGACTCGGATTTCTTCGACTCCTAC AGCATCACTGCCTGCCGCATCGACTGTGAGACGCGCTACCTGGTGGAGAACTGCAACT GCCGCATGGTGCACATGCCAGGTGATGCCCCATACTGTACTCCAGAGCAGTACAAGGA GTGTGCAGATCCTGCTCTGGACTTCCTGGTGGAGAAGGACCAGGAGTACTGCGTGTGT GAAATGCCTTGCAACCTGACCCGCTATGGCAAAGAGCTGTCCATGGTCAAGATCCCCA GCAAAGCCTCAGCCAAGTACCTGGCCAAGAAGTTCAACAAATCTGAGCAATACATAGG GGAGAACATCCTGGTGCTGGACATTTTCTTTGAAGTCCTCAACTATGAGACCATTGAA CAGAAGAAGGCCTATGAGATTGCAGGGCTCCTGGGTGACATCGGGGGCCAGATGGGGC TGTTCATCGGGGCCAGCATCCTCACGGTGCTGGAGCTCTTTGACTACGCCTACGAGGT AGTCATTAAGCACAAGCTGTGCCGACGAGGAAAATGCCAGAAGGAGGCCAAAAGGAGC AGTGCGGACAAGGGCGTGGCCCTCAGCCTGGACGACGTCAAAAGACACAACCCGTGCG AGAGCCTTCGGGGCCACCCTGCCGGGATGACATACGCTGCCAACATCCTACCTCACCA TCCGGCCCGAGGCACGTTCGAGGACTTTACCTGCTGAGCCCCGCAGGCCGCTGAACCA AAGGCCTAGATGGGGAGGACTAGGAGAGCGAGGGGGCCCCCAGCTGCCTCCTCACATC

192

	ORF Start: ATG at 1	ORF Stop: TGA at 1543	
	SEQ ID NO: 104	514 aa	MW at 57221.7kD
NOV30a, CG59241-01 Protein Sequence	MELKAEEEVGGVQPVDLVAFANSCTLHGTNHIFVEGGPGPRQVLWAVAFVLALGAFL CQVGDRVAYYLSYPHVTLLNEVATTELAFPAVTLCNTNAVRLSQLSYPDLLYLAPMLG LDESDDPGVPLAPPGPEAFSGEPFNLHRFYNRSCHRLEDMLLYCSYQGGPCGPHNFSV VFTRYGKCYTFNSGRDGRPRLKTMKGGTGNGLEIMLDIQQDEYLPVWGETDETSFEAG IKVQIHSQDEPPFIDQLGFGVAPGFQTFVACQEQRIYLPPPWGTCKAVTMDSDFFDSY SITACRIDCETRYLVENCNCRMVHMPGDAPYCTPEQYKECADPALDFLVEKDQEYCVC EMPCNLTRYGKELSMVKIPSKASAKYLAKKFNKSEQYIGENILVLDIFFEVLNYETIE QKKAYEIAGLLGDIGGQMGLFIGASILTVLELFDYAYEVVIKHKLCRRGKCQKEAKRS SADKGVALSLDDVKRHNPCESLRGHPAGMTYAANILPHHPARGTFEDFTC		

Further analysis of the NOV30a protein yielded the following properties shown in Table 30B.

Table 30B. Protein Sequence Properties NOV30a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 60 and 61

A search of the NOV30a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 30C.

Table 30C. Geneseq Results for NOV30a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY69178	A rat acid-sensitive cationic channel 1B (rASIC1B) - Rattus sp, 559 aa. [WO200008149-A2, 17-FEB-2000]	1..514 47..559	488/515 (94%) 497/515 (95%)	0.0
AAY03186	Rat Acid sensitive ion channel protein sequence - Rattus sp, 513 aa. [WO9911784-A1, 11-MAR-1999]	1..514 1..513	488/515 (94%) 498/515 (95%)	0.0
AAW68507	Rat acid sensing ionic channel 1B - Rattus sp, 559 aa. [WO9835034-A1, 13-AUG-1998]	1..514 47..559	488/515 (94%) 497/515 (95%)	0.0
AAY69175	A rat acid-sensitive cationic	1..514 1..526	416/527 (78%) 445/527 (83%)	0.0

	526 aa. [WO200008149-A2, 17-FEB-2000]			
AAY03188	Rat Acid sensitive ion channel alpha protein sequence - Rattus sp, 526 aa. [WO9911784-A1, 11-MAR-1999]	1..514 1..526	416/527 (78%) 445/527 (83%)	0.0

In a BLAST search of public sequence databases, the NOV30a protein was found to have homology to the proteins shown in the BLASTP data in Table 30D.

Table 30D. Public BLASTP Results for NOV30a				
Protein Accession Number	Protein/Organism/Length	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q91YB8	ION CHANNEL - Rattus norvegicus (Rat), 559 aa.	1..514 47..559	489/515 (94%) 498/515 (95%)	0.0
O88762	ASIC-BETA - Rattus norvegicus (Rat), 513 aa.	1..514 1..513	488/515 (94%) 498/515 (95%)	0.0
P55926	Amiloride-sensitive brain sodium channel BNaC2 (Amiloride-sensitive cation channel neuronal 2) (Proton gated cation channel ASIC1) - Rattus norvegicus (Rat), 526 aa.	1..514 1..526	416/527 (78%) 445/527 (83%)	0.0
P78348	Amiloride-sensitive brain sodium channel BNaC2 (Amiloride-sensitive cation channel neuronal 2) - Homo sapiens (Human), 574 aa.	1..514 1..574	421/575 (73%) 447/575 (77%)	0.0
Q99NA1	PROTON-GATED CATION CHANNEL SUBUNIT ASIC-BETA2 - Rattus norvegicus (Rat), 425 aa.	175..514 86..425	334/341 (97%) 337/341 (97%)	0.0

PFam analysis predicts that the NOV30a protein contains the domains shown in the

5   Table 30E.

Table 30E. Domain Analysis of NOV30a			
Pfam Domain	**NOV30a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ASC: domain 1 of 2	21..118	34/106 (32%) 79/106 (75%)	1.6e-29
ASC: domain 2 of 2	145..442	133/351 (38%) 281/351 (80%)	2.1e-139

Example 31.

The NOV31 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 31A.

Table 31A. NOV31 Sequence Analysis			
	SEQ ID NO: 105	1949 bp	
NOV31a, CG58602-01 DNA Sequence	TGCCTGGCTATGGCCCGACTGCTCAGGTCTGCAACCTGGGAGCTGTTCCCCTGGAGGG GCTACTGCTCCCAGTCCCTGCAGGGAGAGCTCTGCAGGGACTTCGTAGAGGCTCTGAA GGCCGTGGTGGGCGGCTCCCACGTGTCCACTGCCGCGGTGGTCCGAGAGCAGCACGGG CGCGATGAGTCGGTGCACAGGTGCGAACCTCCTGATGCTGTGGTGTGGCCCCAGAACG TGGAGCAGGTCAGCCGGCTGGCAGCCCTGTGCTATCGCCAAGGTGTGCCCATCATCCC ATTCGGCACCGGCACCGGGCTTGAGGGTGGCGTCTGTGCTGTGCAGGGCGGCGTCTGC GTTAACCTGACGCATATGGACCGAATCCTGGAGCTGAACCAGGAGGACTTCTCTGTGG TGGTGGAGCCAGGTGTCACCCGCAAAGCCCTCAACGCCCACCTGCGGGACAGCGGCCT CTGGTTTCCTCCAGACCCAGGCGCGGACGCCTCTCTCTGTGGCATGGCGGCCACCGGG GCGTCGGGGACCAACGCGGTCCGCTACGGCACCATGCGGGACAACGTGCTCAACCTGG AGGTGGTGCTGCCCGACGGGCGGCTGCTGCACACGGCGGGCCTGAGGCCTCATCACGA TTCCACTGCTGCATTCCCCCACATCAGCCCCACTGAGTGCTTTTCCCAGGGGCCAGGG CCTCATGTCAATTCTCCTCACCCTGCCCCTGAGGCCACAGTGGCCGCCACGTGTGCGT TCCCCAGTGTCCAGGCTGCTGTGGACAGCACTGTACACATCCTCCAGGCTGCAGTGCC CGTAGCCCGCATTGAGTTCCTGGATGAAGTCATGATGGATGCCTGCAACAGGTACAGC AAGCTGAATTGCTTAGTGGCGCCCACACTCTTCCTGGAGTTCCATGGCTCCCAGCAGG CACTGGAGGAGCAGCTGCAGCGCACAGAGGAGATAGTCCAGCAGAACGGAGCCTCTGA CTTCTCCTGGGCCAAGGAGGCCGAGGAGCGCAGCCGGCTTTGGACAGCACGGCACAAT GCCTGGTACGCAGCCCTGGCCACGCGGCCAGGCTGCAAGGGCTACTCCACGGATGTGT GTGTGCCCATCTCCCGGCTGCCGGAGATCGTGGTGCAGACCAAGGAGGATCTGAATGC CTCAGGACTCACAGGAAGCATTGTCGGGCATGTGGGTGACGGCAACTTCCACTGCATC CTGCTGGTCAACCCTGATGACGCCGAGGAACTGGGCAGGGTCAAGGCTTTTGCAGAAC AGCTGGGCAGGCGGGCACTGGCTCTCCACGGAACGTGCACGGGGGAGCATGGCATCGG AATGGGCAAGCGGCAGCTGCTGCAGGAGGAGGTGGGCGCCGTGGGCGTGGAGACCATG CGGCAGCTCAAGGCCGTGCTAGACCCCCAAGGCCTCATGAATCCAGGCAAAGTGCTGT GAAGGGGGTCTGAGCACTTAGCCCACAAGTTCCCTGACTACGGAGCCGGTTCTGGAAC TTTTCTTCATGCCACGGCCCCTGCAAGGAAATAGATGCTGAGGCAGTCTTCCTGCCAG CGAGCCCACTGTATCTGGGCCCAAGGCCAGAGGGCCCAGAGAGAAGCCTGAGCACCGT GTTACCTCCCTGGCCCTCTGGCTGGCCCCAGGAGCCTTTGGTTCAGTAAACGACCCAG GGTGGTTCCCAGCAAAGCTGCTTCCTCTCTGCTCCTACGCATCCTGTCCTGGCGGGAA GAGAGCGTCTGGGTCCATTCAAGACTCTGATGACACCCCTCCCCGAGGCCTCCCACTG CCGGGGTCCCAGGACCCTTCCCCCTTCACCTGGTGACAGGAACACTCCTTTCCTGGTA TGGAACGTGAGCTCCCGTGACATGATGATAGGTCTTCTCCTTGGGGCCTCCCCCAATA AATCTGTAATAAACCTGAAACCCACCTACAGCTAA
	ORF Start: ATG at 10	ORF Stop: TGA at 1450	
	SEQ ID NO: 106	480 aa	MW at 51629.1kD
NOV31a, CG58602-01 Protein Sequence	MARLLRSATWELFPWRGYCSQSLQGELCRDFVEALKAVVGGSHVSTAAVVREQHGRDE SVHRCEPPDAVVWPQNVEQVSRLAALCYRQGVPIIPFGTGTGLEGGVCAVQGGVCVNL THMDRILELNQEDFSVVVEPGVTRKALNAHLRDSGLWFPPDPGADASLCGMAATGASG TNAVRYGTMRDNVLNLEVVLPDGRLLHTAGRGLITDSTAAFPHISPTECFSQGPGPHV NSPHPAPEATVAATCAFPSVQAAVDSTVHILQAAVPVARIEFLDEVMMDACNRYSKLN CLVAPTLFLEFHGSQQALEEQLQRTEEIVQQNGASDFSWAKEAEERSRLWTARHNAWY		

195

	AALATRPGCKGYSTDVCVPISRLPEIVVQTKEDLNASGLTGSIVGHVGDGNFHCILLV
·	NPDDAEELGRVKAFAEQLGRRALALHGTCTGEHGIGMGKRQLLQEEVGAVGVETMRQL
	KAVLDPQGLMNPGKVL

Further analysis of the NOV31a protein yielded the following properties shown in Table 31B.

Table 31B. Protein Sequence Properties NOV31a	
PSort analysis:	0.6574 probability located in mitochondrial matrix space; 0.3502 probability located in mitochondrial inner membrane; 0.3502 probability located in mitochondrial intermembrane space; 0.3502 probability located in mitochondrial outer membrane
SignalP analysis:	Likely cleavage site between residues 20 and 21

A search of the NOV31a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 31C.

Table 31C. Geneseq Results for NOV31a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB10446	Human cDNA SEQ ID NO: 754 - Homo sapiens, 115 aa. [WO200154474-A2, 02-AUG-2001]	1..96 15..110	91/96 (94%) 92/96 (95%)	8e-49
AAE09597	Human gene 5 encoded novel protein HDPMT22, SEQ ID NO:33 - Homo sapiens, 115 aa. [WO200155311-A2, 02-AUG-2001]	1..96 15..110	91/96 (94%) 92/96 (95%)	8e-49
AAM52368	GIP12-C4 protein - Arabidopsis thaliana, 159 aa. [FR2806095-A1, 14-SEP-2001]	66..203 3..140	69/138 (50%) 98/138 (71%)	9e-34
AAG92286	C glutamicum protein fragment SEQ ID NO: 6040 - Corynebacterium glutamicum, 948 aa. [EP1108790-A2, 20-JUN-2001]	46..477 25..502	108/486 (22%) 186/486 (38%)	2e-22
AAB79309	Corynebacterium glutamicum SMP protein sequence SEQ ID NO:134 - Corynebacterium glutamicum, 945 aa. [WO200100844-A2, 04-JAN-2001]	46..477 22..499	108/486 (22%) 186/486 (38%)	2e-22

In a BLAST search of public sequence databases, the NOV31a protein was found to have homology to the proteins shown in the BLASTP data in Table 31D.

Table 31D. Public BLASTP Results for NOV31a				
Protein Accession Number	Protein/Organism/Length	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D635	4733401P21RIK PROTEIN - Mus musculus (Mouse), 481 aa.	1..480 1..481	394/483 (81%) 423/483 (87%)	0.0
Q19965	F32D8.4 PROTEIN - Caenorhabditis elegans, 912 aa.	20..480 445..909	221/466 (47%) 307/466 (65%)	e-121
CAD16371	PUTATIVE D-LACTATE DEHYDROGENASE (CYTOCHROME) OXIDOREDUCTASE PROTEIN (EC 1.1.2.4) - Ralstonia solanacearum (Pseudomonas solanacearum), 472 aa.	32..479 20..469	226/454 (49%) 300/454 (65%)	e-119
A89201	protein F32D8.4 [imported] - Caenorhabditis elegans, 870 aa.	30..480 399..867	214/469 (45%) 296/469 (62%)	e-115
AAL51780	D-LACTATE DEHYDROGENASE (CYTOCHROME) (EC 1.1.2.4) - Brucella melitensis, 468 aa.	41..480 28..467	209/444 (47%) 286/444 (64%)	e-114

PFam analysis predicts that the NOV31a protein contains the domains shown in the Table 31E.

Table 31E. Domain Analysis of NOV31a			
Pfam Domain	NOV31a Match Region	Identities/ Similarities for the Matched Region	Expect Value
FAD_binding_4: domain 1 of 1	33..214	70/208 (34%) 154/208 (74%)	3.7e-56
FAD-oxidase_C: domain 1 of 1	206..479	91/307 (30%) 210/307 (68%)	1.3e-58

Example 32.

The NOV32 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 32A.

197

Table 32A. NOV32 Sequence Analysis		
	SEQ ID NO: 107	698 bp
NOV32a, CG58468-01 DNA Sequence	CTCCTTCCTGTGCTCTTTATATGGACCAACAACTTCTCGTCCTTGGGGTCTCTGTGCA AATATCAATTTTTTCACATTATCTTTTCTCCACAGACATGAGAGGGAAGGCATTTATT TTCCCTCAAGAATCAGCTACAGTCTATGTGTCCCTGATCCCCAAGGTGAAGAAGCCCC TGAAGAACTTCAAGCTTTGCCTGAAAACCTTCACAGACTTCACCTGCCCTTATAGCCT CTTCTACAGCACTCGGTCCCAGGACAATGAGCTGCTTCTCCTTGTCAACAAAATGGGA ATGTATCTGCTGCACATTGGAAATGCTGCGGTCACTTTCAATGGCCCCACCCCCTGCC CTCGATCTCCTTATGCTTCGACCCATGTCAATGTGAGCTGGGAGTCTGCCTCTGGAAT TGCTACACTCTGGGCAAATGGGAAGCTGGTGGGGAGGAAGGGTGTGTGGAAGGGGTAC TCTGTGGGAGAAGAGGCTAAGATCATCCTGGGACAAGAGCAGGATTCCTTTGGGGGAC ATTTTGATGAAAATCAATCCTTTGTTGGGGTGATATGGGATGTGTTTTTGTGGGATCA TGTGCTCCCTCCAAAGGAGATGTGTGACTCCTGTTACAGCGGCAGCCTCCTGAATCGG CATACCCTGACTTATGAAGATAATGGCTATGTGGTAACTAAGCCCAAGGTGTGGGCTT AA
	ORF Start: ATG at 21	ORF Stop: TAA at 696
	SEQ ID NO: 108	225 aa	MW at 25265.8kD
NOV32a, CG58468-01 Protein Sequence	MDQQLLVLGVSVQISIFSHYLFSTDMRGKAFIFPQESATVYVSLIPKVKKPLKNFKLC LKTFTDFTCPYSLFYSTRSQDNELLLLVNKMGMYLLHIGNAAVTFNGPTPCPRSPYAS THVNVSWESASGIATLWANGKLVGRKGVWKGYSVGEEAKIILGQEQDSFGGHFDENQS FVGVIWDVFLWDHVLPPKEMCDSCYSGSLLNRHTLTYEDNGYVVTKPKVWA	

Further analysis of the NOV32a protein yielded the following properties shown in Table 32B.

Table 32B. Protein Sequence Properties NOV32a	
PSort analysis:	0.5500 probability located in endoplasmic reticulum (membrane); 0.3200 probability located in microbody (peroxisome); 0.2368 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV32a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 32C.

Table 32C. Geneseq Results for NOV32a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR74763	Sermun amyloid P component, promoter sapm - Homo sapiens, 204 aa. [WO9505394-A, 23-FEB-1995]	24..224 2..203	98/207 (47%) 136/207 (65%)	4e-48

198

AAR29923	SAP - Homo sapiens, 223 aa. [WO9221364-A, 10-DEC-1992]	7..224 5..222	101/224 (45%) 143/224 (63%)	3e-47
AAR29922	CRP - Homo sapiens, 225 aa. [WO9221364-A, 10-DEC-1992]	14..224 11..224	100/218 (45%) 132/218 (59%)	2e-43
AAR74769	Female hamster protein, 1fhp - Cricetus cricetus, 210 aa. [WO9505394-A, 23-FEB-1995]	24..222 1..199	95/206 (46%) 132/206 (63%)	6e-43
AAY76844	Human C reactive protein (CRP) sequence - Homo sapiens, 206 aa. [JP2000014388-A, 18-JAN-2000]	24..224 2..205	98/208 (47%) 128/208 (61%)	1e-42

In a BLAST search of public sequence databases, the NOV32a protein was found to have homology to the proteins shown in the BLASTP data in Table 32D.

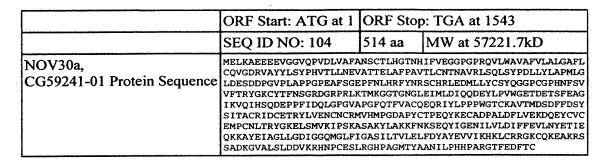| Table 32D. Public BLASTP Results for NOV32a				
Protein Accession Number	Protein/Organism/Length	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D8J8	1810030J14RIK PROTEIN - Mus musculus (Mouse), 219 aa.	6..224 4..218	130/220 (59%) 166/220 (75%)	5e-72
Q9D8V2	1810030J14RIK PROTEIN - Mus musculus (Mouse), 200 aa.	6..190 20..200	110/186 (59%) 139/186 (74%)	2e-58
Q63913	SERUM AMYLOID P - Cricetulus migratorius (Armenian hamster), 223 aa.	1..224 1..222	109/231 (47%) 152/231 (65%)	4e-51
P23680	Serum amyloid P-component precursor (SAP) - Rattus norvegicus (Rat), 228 aa.	6..224 4..223	105/224 (46%) 145/224 (63%)	7e-50
P15697	Female protein precursor (FP) (Serum amyloid P-component) - Cricetulus migratorius (Armenian hamster), 231 aa.	1..222 1..220	108/229 (47%) 151/229 (65%)	7e-50

PFam analysis predicts that the NOV32a protein contains the domains shown in the

5   Table 32E.

<table>
<tr><th colspan="4">Table 32E. Domain Analysis of NOV32a</th></tr>
</table>

Pfam Domain	NOV32a Match Region	Identities/ Similarities for the Matched Region	Expect Value
pentaxin: domain 1 of 1	29..221	103/214 (48%) 156/214 (73%)	8e-76

Example 33.

The NOV33 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 33A.

<table>
<tr><th colspan="2">Table 33A. NOV33 Sequence Analysis</th></tr>
<tr><td></td><td>SEQ ID NO: 109   3350 bp</td></tr>
<tr><td>NOV33a,
CG58183-01 DNA Sequence</td><td>

TAATGAGGAGACTGAGTTTGTGGTGGCTGCTGAGCAGGGTCTGTCTGCTGTTGCCGCC
GCCCTGCGCACTGGTGCTGGCCGGGGTGCCCAGCTCCTCCTCGCACCCGCAGCCCTGC
CAGATCCTCAAGCGCATCGGGCACGCGGTGAGGGTGGGCGCGGTGCACTTGCAGCCCT
GGACCACCGCCCCCCGCGCGGCCAGCCGCGCTCCGGACGACAGCCGAGCAGGAGCCCA
GAGGGATGAGCCGGAGCCAGGGACTAGGCGGTCCCCGGCGCCCTCGCCGGGCGCACGC
TGGTTGGGGAGCACCCTGCATGGCCGGGGGCCGCCGGGCTCCCGTAAGCCCGGGGAGG
GCGCCAGGGCGGAGGCCCTGTGGCCACGGGACGCCCTCCTATTTGCCGTGGACAACCT
GAACCGCGTGGAAGGGCTGCTACCCTACAACCTGTCTTTGGAAGTAGTGATGGCCATC
GAGGCAGGCCTGGGCGATCTGCCACTTTTGCCCTTCTCCTCCCCTAGTTCGCCATGGA
GCAGTGACCCTTTCTCCTTCCTGCAAAGTGTGTGCCATACCGTGGTGGTGCAAGGGGT
GTCGGCGCTGCTCGCCTTCCCCCAGAGCCAGGGCGAAATGATGGAGCTCGACTTGGTC
AGCTTAGTCCTGCACATTCCAGTGATCAGCATCGTGCGCCACGAGTTTCCACGGGAGA
GTCAGAATCCCCTTCACCTACAACTGAGTTTAGAAAATTCATTAAGTTCTGATGCTGA
TGTCACTGTCTCAATCCTGACCATGAACAACTGGTACAATTTTAGCTTGTTGCTGTGC
CAGGAAGACTGGAACATCACCGACTTCCTCCTCCTTACCCAGAATAATTCCAAGTTCC
ACCTTGGTTCTATCATCAACATCACCGCTAACCTCCCCTCCACCCAGGACCTCTTGAG
CTTCCTACAGATCCAGCTTGAGAGTATTAAGAACAGCACACCCACAGTGGTGATGTTT
GGCTGCGACATGGAAAGTATCCGGCGGATTTTCGAAATTACAACCCAGTTTGGGGTCA
TGCCCCCTGAACTTCGTTGGGTGCTGGGAGATTCCCAGAATGTGGAGGAACTGAGGAC
AGAGGGTCTGCCCTTAGGGCTCATTGCTCATGGAAAAACAACACAGTCTGTCTTTGAG
CACTACGTACAAGATGCTATGGAGCTGGTCGCAAGAGCTGTAGCCACAGCCACCATGA
TCCAACCAGAACTTGCTCTCATTCCCAGCACGATGAACTGCATGGAGGTGGAAACTAC
AAATCTCACTTCAGGACAATATTTATCAAGGTTTCTAGCCAATACCACTTTCAGAGGC
CTCAGTGGTTCCATCAGAGTAAAAGGTTCCACCATCGTCAGCTCAGAAAACAACTTTT
TCATCTGGAATCTTCAACATGACCCCATGGGAAAGCCAATGTGGACCCGCTTGGGCAG
CTGGCAGGGGGGAAAGATTGTCATGGACTATGGAATATGGCCAGAGCAGGCCCAGAGA
CACAAAACCCACTTCCAACATCCAAGTAAGCTACACTTGAGAGTGGTTACCCTGATTG
AGCATCCTTTTGTCTTCACAAGGGAGGTAGATGATGAAGGCTTGTGCCCTGCTGGCCA
ACTCTGTCTAGACCCCATGACTAATGACTCTTCCACATTGGACAGCCTTTTTAGCAGC
CTCCATAGCAGTAATGATACAGTGCCCATTAAATTCAAGAAGTGCTGCTATGGATATT
GCATTGATCTGCTGGAAAAGATAGCAGAAGACATGAACTTTGACTTCGACCTCTATAT
TGTAGGGGATGGAAAGTATGGAGCATGGAAAAATGGGCACTGGACTGGGCTAGTGGGT
GATCTCCTGAGAGGGACTGCCCACATGGCAGTCACTTCCTTTAGCATCAATACTGCAC
GGAGCCAGGTGATAGATTTCACCAGCCCTTTCTTCTCCACCAGCTTGGGCATCTTAGT
GAGGACCCGAGATACAGCAGCTCCCATTGGAGCCTTCATGTGGCCACTCCACTGGACA
ATGTGGCTGGGGATTTTTGTGGCTCTGCACATCACTGCCGTCTTCCTCACTCTGTATG
AATGGAAGAGTCCATTTGGTTTGACTTCCAAGGGGCGAAATAGAAGTAAAGTCTTCTC
CTTTTCTTCAGCCTTGAACATCTGTTATGCCCTCTTGTTTGGCAGAACAGTGGCCATC
AAACCTCCAAAATGTTGGACTGGAAGGTTTCTAATGAACCTTTGGGCCATTTTCTGTA
TGTTTTGCCTTTCCACATACACGGCAAACTTGGCTGCTGTCATGGTAGGTGAGAAGAT
CTATGAAGAGCTTTCTGGAATACATGACCCCAAGTTACATCATCCTTCCCAAGGATTC
CGCTTTGGAACTGTCCGAGAAAGCAGTGCTGAAGATTATGTGAGACAAAGTTTCCCAG
AGATGCATGAATATATGAGAAGGTACAATGTTCCAGCCACCCCTGATGGAGTGGAGTA
TCTGAAGAATGATCCAGAGAAACTAGACGCCTTCATCATGGACAAAGCCCTTCTGGAT
TATGAAGTGTCAATAGATGCTGACTGCAAACTTCTCACTGTGGGGAAGCCATTTGCCA
TAGAAGGTTACGGCATTGGCCTCCCCACCCAACTCTCCATTGACCGCCAACATATCCGA
GCTAATCAGTCAATACAAGTCACATGGGTTTATGGATATGCTCCATGACAAGTGGTAC

</td></tr>
</table>

200

	AGGGTGGTTCCCTGTGGCAAGAGAAGTTTTGCTGTCACGGAGACTTTGCAAATGGGCA TCAAACACTTCTCTGGGCTCTTTGTGCTGCTGTGCATTGGATTTGGTCTGTCCATTTT GACCACCATTGGTGAGCACATAGTATACAGGCTGCTGCTACCACGAATCAAAAACAAA TCCAAGCTGCAATACTGGCTCCACACCAGCCAGAGATTACACAGAGCAATAAATACAT CATTTATAGAGGAAAAGCAGCAGCATTTCAAGACCAAACGTGTGGAAAAGAGATCTAA TGTGGGACCCCGTCAGCTTACCGTATGGAATACTTCCAATCTGAGTCATGACAACCGA CGGAAATACATCTTTAGTGATGAGGAAGGACAAAACCAGCTGGGCATCCGGATCCACC AGGACATCCCCCTCCCTCCAAGGAGAAGAGAGCTCCCTGCCTTGCGGACCACCAATGG GAAAGCAGACTCCCTAAATGTATCTCGGAACTCAGTGATGCAGGAACTCTCAGAGCTC GAGAAGCAGATTCAGGTGATCCGTCAGGAGCTGCAGCTGGCTGTGAGCAGGAAAACGG AGCTGGAGGAGTATCAAAGGACAAGTCGGACTTGTGAGTCC**TAG**

	ORF Start: ATG at 3	ORF Stop: TAG at 3348	
	SEQ ID NO: 110	1115 aa	MW at 125453.7kD
NOV33a, CG58183-01 Protein Sequence	MRRLSLWWLLSRVCLLLPPPCALVLAGVPSSSSHPQPCQILKRIGHAVRVGAVHLQPW TTAPRAASRAPDDSRAGAQRDEPEPGTRRSPAPSPGARWLGSTLHGRGPPGSRKPGEG ARAEALWPRDALLFAVDNLNRVEGLLPYNLSLEVVMAIEAGLGDLPLLPFSSPSSPWS SDPFSFLQSVCHTVVVQGVSALLAFPQSQGEMMELDLVSLVLHIPVISIVRHEFPRES QNPLHLQLSLENSLSSDADVTVSILTMNNWYNFSLLLCQEDWNITDFLLLTQNNSKFH LGSIINITANLPSTQDLLSFLQIQLESIKNSTPTVVMFGCDMESIRRIFEITTQFGVM PPELRWVLGDSQNVEELRTEGLPLGLIAHGKTTQSVFEHYVQDAMELVARAVATATMI QPELALIPSTMNCMEVETTNLTSGQYLSRFLANTTFRGLSGSIRVKGSTIVSSENNFF IWNLQHDPMGKPMWTRLGSWQGGKIVMDYGIWPEQAQRHKTHFQHPSKLHLRVVTLIE HPFVFTREVDDEGLCPAGQLCLDPMTNDSSTLDSLFSSLHSSNDTVPIKFKKCCYGYC IDLLEKIAEDMNFDFDLYIVGDGKYGAWKNGHWTGLVGDLLRGTAHMAVTSFSINTAR SQVIDFTSPFFSTSLGILVRTRDTAAPIGAFMWPLHWTMWLGIFVALHITAVFLTLYE WKSPFGLTSKGRNRSKVFSFSSALNICYALLFGRTVAIKPPKCWTGRFLMNLWAIFCM FCLSTYTANLAAVMVGEKIYEELSGIHDPKLHHPSQGFRFGTVRESSAEDYVRQSFPE MHEYMRRYNVPATPDGVEYLKNDPEKLDAFIMDKALLDYEVSIDADCKLLTVGKPFAI EGYGIGLPPNSPLTANISELISQYKSHGFMDMLHDKWYRVVPCGKRSFAVTETLQMGI KHFSGLFVLLCIGFGLSILTTIGEHIVYRLLLPRIKNKSKLQYWLHTSQRLHRAINTS FIEEKQQHFKTKRVEKRSNVGPRQLTVWNTSNLSHDNRRKYIFSDEEGQNQLGIRIHQ DIPLPPRRRELPALRTTNGKADSLNVSRNSVMQELSELEKQIQVIRQELQLAVSRKTE LEEYQRTSRTCES

Further analysis of the NOV33a protein yielded the following properties shown in Table 33B.

Table 33B. Protein Sequence Properties NOV33a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 34 and 35

A search of the NOV33a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 33C.

Table 33C. Geneseq Results for NOV33a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU02199	Human glutamate receptor-like protein, MEM4 - Homo sapiens, 1043 aa. [WO200144473-A2, 21-JUN-2001]	95..1103 6..1007	508/1047 (48%) 680/1047 (64%)	0.0
AAB42494	Human ORFX ORF2258 polypeptide sequence SEQ ID NO:4516 - Homo sapiens, 901 aa. [WO200058473-A2, 05-OCT-2000]	95..985 6..885	484/912 (53%) 635/912 (69%)	0.0
AAU02198	Human glutamate receptor-like protein, MEM3 - Homo sapiens, 971 aa. [WO200144473-A2, 21-JUN-2001]	532..1103 362..935	361/579 (62%) 448/579 (77%)	0.0
AAU02197	Human glutamate receptor-like protein, MEM2 - Homo sapiens, 965 aa. [WO200144473-A2, 21-JUN-2001]	532..1103 362..929	352/579 (60%) 437/579 (74%)	0.0
AAR44192	Rat NMDA receptor subunit, NR2A - Rattus rattus, 1464 aa. [DE4216321-A, 18-NOV-1993]	175..1023 77..911	245/873 (28%) 425/873 (48%)	2e-83

In a BLAST search of public sequence databases, the NOV33a protein was found to have homology to the proteins shown in the BLASTP data in Table 33D.

Table 33D. Public BLASTP Results for NOV33a				
Protein Accession Number	Protein/Organism/Length	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAL40734	N-METHYL-D-ASPARTATE RECEPTOR 3A - Homo sapiens (Human), 1115 aa.	1..1115 1..1115	1110/1115 (99%) 1112/1115 (99%)	0.0
Q62800	IONOTROPIC GLUTAMATE RECEPTOR - Rattus norvegicus (Rat), 1115 aa.	1..1115 1..1115	1032/1115 (92%) 1083/1115 (96%)	0.0

202

Q9R1M7	N-METHYL-D-ASPARTATE RECEPTOR SPLICE VARIANT NR3A-2 - Rattus norvegicus (Rat), 1135 aa.	1..1115 1..1135	1032/1135 (90%) 1083/1135 (94%)	0.0
CAC69380	SEQUENCE 7 FROM PATENT WO0144473 - Homo sapiens (Human), 1043 aa.	95..1103 6..1007	508/1047 (48%) 680/1047 (64%)	0.0
Q91ZU9	NMDA-TYPE GLUTAMATE RECEPTOR SUBUNIT NR3B PRECURSOR - Mus musculus (Mouse), 1003 aa.	112..1103 34..980	510/1001 (50%) 669/1001 (65%)	0.0

PFam analysis predicts that the NOV33a protein contains the domains shown in the Table 33E.

Table 33E. Domain Analysis of NOV33a			
Pfam Domain	NOV33a Match Region	Identities/ Similarities for the Matched Region	Expect Value
lig_chan: domain 1 of 1	674..952	81/323 (25%) 232/323 (72%)	4e-95

Example 34.

The NOV34 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 34A.

Table 34A. NOV34 Sequence Analysis	
	SEQ ID NO: 111 1253 bp
NOV34a, CG59315-01 DNA Sequence	CCCAGCGCCATGGGGGAGTGGGCGTTCCTGGGCTCGCTGCTGGACGCCGTGCAGCTGC AGTCGCCGCTCGTGGGCCGCCTCTGGCTGGTGGTCATGCTGATCTTCCGCATCCTGGT GCTGGCCACGGTGGGCGGCGCCGTGTTCGAGGACGAGCAAGAGGAGTTCGTGTGCAAC ACGCTGCAGCCGGGCTGTCGCCAGACCTGCTACGACCGCGCCTTCCCGGTCTCCCACT ACCGCTTCTGGCTCTTCCACATCCTGCTGCTCTCGGCGCCCCCGGTGCTGTTCGTCGT CTACTCCATGCACCGGGCAGGCAAGGAGGCGGGCGGCGCTGAGGCGGCGGCGCAGTGC GCCCCCGGACTGCCCGAGGCCCAGTGCGCGCCGTGCGCCCTGCGCGCCCGCCGCGCGC GCCGCTGCTACCTGCTGAGCGTGGCGCTGCGCCTGCTGGCCGAGCTGACCTTCCTGGG CGGCCAGGCGCTGCTCTACGGCTTCCGCGTGGCCCCGCACTTCGCGTGCGCCGGTCCG CCCTGCCCGCACACGGTCGACTGCTTCGTGAGCCGGCCCACCGAGAAGACCGTCTTCG TGCTCTTCTATTTCGCGGTGGGGCTGCTGTCGGCGCTGCTCAGCGTAGCCGAGCTGGG CCACCTGCTCTGGAAGGGCCGCCCGCGCGCCGGGGAGCGTGACAACCGCTGCAACCGT GCACACGAAGAGGCGCAGAAGCTGCTCCCGCCGCCGCCGCCGCCACCTCCGCCACCGG CCCTGCCCTCCCGGCGCCCCGGCCCCGAGCCGTGCGCCCCGCCGGCCTATGCGCACCC GGCGCCGGCCAGCCTCCGCGAGTGCGGCAGCGGCCGCGGCAGGAATGCGCCAATGGCT CCCAGATGTGGACGCCACCGCTTAACCCCTTACCCCCCAGCCGCGCTCCCCCAAGGGC CTTCCAGCCTGAGCCCCGCCAACAGCAGGGAGCTCTGCCCAGGTGAGAACCAGCCCAG GACTGGAGTCAGCGCCAGCCCGCCCCTAGTGCCCACGGACACCTCCCAACCTAGATCC TACCTGTCTTCCTTCCTTGAGGCTGGAGGGGAAGGCTCATGGACACAAGAATGCAAGC ATGCATGCACACAGCTACACTGCCTCCCATCCCCTCCCGCCGACGCTGCCAGGGTGCC CCTCCCTCGCTCCCCATCCTGGCAGGGCGGGCGGCGCAGAGCGCTCCACTCCGGATTC CCCACGCCCCCGAGCCGTTCGCAGGCTCGCACAAG

203

	ORF Start: ATG at 10	ORF Stop: AG at 1252
	SEQ ID NO: 112	414 aa	MW at 44773.0kD
NOV34a, CG59315-01 Protein Sequence	MGEWAFLGSLLDAVQLQSPLVGRLWLVVMLIFRILVLATVGGAVFEDEQEEFVCNTLQ PGCRQTCYDRAFPVSHYRFWLFHILLLSAPPVLFVVYSMHRAGKEAGGAEAAAQCAPG LPEAQCAPCALRARRARRCYLLSVALRLLAELTFLGGQALLYGFRVAPHFACAGPPCP HTVDCFVSRPTEKTVFVLFYFAVGLLSALLSVAELGHLLWKGRPRAGERDNRCNRAHE EAQKLLPPPPPPPPPPALPSRRPGPEPCAPPAYAHPAPASLRECGSGRGRNAPMAPRC GRHRLTPYPPAALPQGPSSLSPANSRELCPGENQPRTGVSASPPLVPTDTSQPRSYLS SFLEAGGEGSWTQECKHACTQLHCLPSPPADAARVPLPRSPSWQGGRRRALHSGFPTP PSRSQART

Further analysis of the NOV34a protein yielded the following properties shown in Table 34B.

Table 34B. Protein Sequence Properties NOV34a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 39 and 40

A search of the NOV34a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 34C.

Table 34C. Geneseq Results for NOV34a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW49009	Mouse alpha 3 connexin protein - Mus sp, 417 aa. [WO9830677-A1, 16-JUL-1998]	1..296 1..327	121/334 (36%) 169/334 (50%)	5e-52
AAW23968	Connexin protein Cx40 - Homo sapiens, 358 aa. [WO9802150-A1, 22-JAN-1998]	1..215 1..232	93/233 (39%) 133/233 (56%)	9e-46
AAW23970	Connexin protein Cx45 - Homo sapiens, 396 aa. [WO9802150-A1, 22-JAN-1998]	4..212 3..253	93/252 (36%) 137/252 (53%)	3e-43
AAW23969	Connexin protein Cx43 - Homo sapiens, 382 aa. [WO9802150-A1, 22-JAN-1998]	1..216 1..235	86/235 (36%) 130/235 (54%)	1e-42
AAM93194	Human polypeptide, SEQ ID NO:	7..384 7..360	129/409 (31%) 169/409 (40%)	8e-38

[EP1130094-A2, 05-SEP-2001]			

In a BLAST search of public sequence databases, the NOV34a protein was found to have homology to the proteins shown in the BLASTP data in Table 34D.

Table 34D. Public BLASTP Results for NOV34a				
Protein Accession Number	Protein/Organism/Length	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q91YD1	CONNEXIN30.2 - Mus musculus (Mouse), 278 aa.	1..283 1..265	228/283 (80%) 240/283 (84%)	e-129
I46053	connexin44 - bovine, 402 aa.	1..397 1..396	151/418 (36%) 207/418 (49%)	1e-62
P41987	Gap junction alpha-3 protein (Connexin 44) (Cx44) - Bos taurus (Bovine), 401 aa.	2..397 1..395	150/417 (35%) 206/417 (48%)	4e-62
AAA50954	CONNEXIN44 - Bos taurus (Bovine), 407 aa.	1..398 1..402	154/429 (35%) 214/429 (48%)	1e-60
Q9TU17	GAP JUNCTION PROTEIN (CONNEXIN) - Ovis aries (Sheep), 413 aa.	1..398 1..408	147/415 (35%) 204/415 (48%)	1e-60

PFam analysis predicts that the NOV34a protein contains the domains shown in the Table 34E.

Table 34E. Domain Analysis of NOV34a			
Pfam Domain	NOV34a Match Region	Identities/ Similarities for the Matched Region	Expect Value
DUF26: domain 1 of 1	107..152	12/56 (21%) 27/56 (48%)	1.4
connexin: domain 1 of 1	1..212	101/247 (41%) 150/247 (61%)	6.5e-75

Example 35.

The NOV35 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 35A.

205

Table 35A. NOV35 Sequence Analysis		
	SEQ ID NO: 113	724 bp
NOV35a, CG59203-01 DNA Sequence	TAAATTCGCGGCCGCGTCGACCTTCCGCAGACTCAACTGAGAAGTCAGCCTCTGCGGC AGGCACCAGGAATCTGCCTTTTCAGTTCTGTCTCCGGCAGGCTTTGAGGATGAAGGCT GCGGGCATTCTGACCCTCATTGGCTGCCTGGTCACAGGCGCCGAGTCCAAAATCTACA CTCGTTGCAAACTGGCAAAAATATTCTCGAGGGCTGGCCTGGACAATTACTGGGGCTT CAGCCTTGGAAACTGGATCTGCATGGCGTATTATGAGAGCGGCTACAACACCACAGCC CAGACGGTCCTGGATGACGGCAGCATCGACTACGGCATCTTCCAGATCAACAGCTTCG CGTGGTGCAGACGCGGGAAAGCTGAAGGAGAACAACCACTGCCACGTCGCCTGCTCAGC CTTGGTCACTGATGACCTCACAGATGCGATTATCTGTGCCAAGAAAATTGTTAAAGAG ACACAAGGAATGAACTATTGGCAAGGCTGGAAGAAACACTGTGAGGGGAGAGACCTGT CCGACTGGAAAAAAGACTGTGAGGTTTCCTAAACTGGAACTGGACCCAGGATGCTTTG CAGCAACGCCCTAGGGTTTGCAGTGAATGTCCAAATGCCTGTGTCATCTTGTCCCGTT TCCTCCCAATATTCCTTCTCAAACTTGGAGAGGGAAAATTAAGCTATACTTTTAAGAA AATAAATATTTCCATTTAAATGTCAAAA
	ORF Start: ATG at 108	ORF Stop: TAA at 552
	SEQ ID NO: 114	148 aa	MW at 16655.9kD
NOV35a, CG59203-01 Protein Sequence	MKAAGILTLIGCLVTGAESKIYTRCKLAKIFSRAGLDNYWGFSLGNWICMAYYESGYN TTAQTVLDDGSIDYGIFQINSFAWCRRGKLKENNHCHVACSALVTDDLTDAIICAKKI VKETQGMNYWQGWKKHCEGRDLSDWKKDCEVS	
	SEQ ID NO: 115	453 bp
NOV35b, CG59203-02 DNA Sequence	CATTCTGACCCTCATTGGCTGCCTGGTCACAGGCGCCGAGTCCAAAATCTACACTCGT TGCAAACTGGCAAAAATATTCTCGAGGGCTGGCCTGGACAATTACTGGGGCTTCAGCC TTGGAAACTGGATCTGCATGGCGTATTATGAGAGCGGCTACAACACCACAGCCCAGAC GGTCCTGGATGACGGCAGCATCGACTACGGCATCTTCCAGATCAACAGCTTCGCGTGG TGCAGACGCGGGAAAGCTGAAGGAGAACAACCACTGCCACGTCGCCTGCTCAGCCTTGG TCACTGATGACCTCACAGATGCAATTATCTGTGCCAGGAAAATTGTTAAAGAGACACA AGGAATGAATTATTGGCAAGGCTGGAAGAAACATTGTGAGGGCAGAGACCTGTCCGAC TGGAAAAAAGGCTGTGAGGTTTCCTAAACTGGAACTGGACCCAGGAT
	ORF Start: ATG at 134	ORF Stop: TAA at 431
	SEQ ID NO: 116	99 aa	MW at 11288.6kD
NOV35b, CG59203-02 Protein Sequence	MAYYESGYNTTAQTVLDDGSIDYGIFQINSFAWCRRGKLKENNHCHVACSALVTDDLT DAIICARKIVKETQGMNYWQGWKKHCEGRDLSDWKKGCEVS	

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 35B.

Table 35B. Comparison of NOV35a against NOV35b.		
Protein Sequence	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV35b	50..148 1..99	97/99 (97%) 98/99 (98%)

Further analysis of the NOV35a protein yielded the following properties shown in Table 35C.

Table 35C. Protein Sequence Properties NOV35a	
Psort analysis:	0.3700 probability located in outside; 0.1697 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 20 and 21

A search of the NOV35a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 35D.

Table 35D. Geneseq Results for NOV35a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY57399	Human lysoenzyme LYC2 polypeptide - Homo sapiens, 148 aa. [WO200012722-A1, 09-MAR-2000]	1..148 1..148	143/148 (96%) 147/148 (98%)	3e-86
AAU29169	Human PRO polypeptide sequence #146 - Homo sapiens, 148 aa. [WO200168848-A2, 20-SEP-2001]	1..148 1..148	143/148 (96%) 146/148 (98%)	6e-86
AAB66145	Protein of the invention #57 - Unidentified, 148 aa. [WO200078961-A1, 28-DEC-2000]	1..148 1..148	143/148 (96%) 146/148 (98%)	6e-86
AAY99396	Human PRO1278 (UNQ648) amino acid sequence SEQ ID NO:203 - Homo sapiens, 148 aa. [WO200012708-A2, 09-MAR-2000]	1..148 1..148	143/148 (96%) 146/148 (98%)	6e-86
AAY71109	Human Hydrolase protein-7 (HYDRL-7) - Homo sapiens, 194 aa. [WO200028045-A2, 18-MAY-2000]	1..148 47..194	142/148 (95%) 146/148 (97%)	1e-85

5   In a BLAST search of public sequence databases, the NOV35a protein was found to have homology to the proteins shown in the BLASTP data in Table 35E.

207

Table 35E. Public BLASTP Results for NOV35a				
Protein Accession Number	Protein/Organism/Length	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96LF2	BA14C22.1 (NOVEL PROTEIN SIMILAR TO LYSOZYME) - Homo sapiens (Human), 148 aa.	1..148 1..148	148/148 (100%) 148/148 (100%)	7e-88
Q9H1R9	BA534G20.1.1 (NOVEL PROTEIN SIMILAR TO LYSOZYME C-1 (1,4-BETA-N-ACYLMURAMIDASE C, EC 3.2.1.17) (ISOFORM 1)) - Homo sapiens (Human), 148 aa.	1..148 1..148	144/148 (97%) 147/148 (99%)	4e-86
AAH21730	HYPOTHETICAL 21.6 KDA PROTEIN - Homo sapiens (Human), 194 aa.	1..148 47..194	143/148 (96%) 146/148 (98%)	2e-85
Q9CPX3	1700038F02RIK PROTEIN - Mus musculus (Mouse), 148 aa.	1..148 1..148	110/148 (74%) 127/148 (85%)	3e-66
Q9H1R8	BA534G20.1.2 (NOVEL PROTEIN SIMILAR TO LYSOZYME C-1 (1,4-BETA-N-ACYLMURAMIDASE C, EC 3.2.1.17) (ISOFORM 2)) - Homo sapiens (Human), 106 aa (fragment).	20..125 1..106	104/106 (98%) 106/106 (99%)	1e-59

PFam analysis predicts that the NOV35a protein contains the domains shown in the Table 35F.

Table 35F. Domain Analysis of NOV35a			
Pfam Domain	NOV35a Match Region	Identities/ Similarities for the Matched Region	Expect Value
lys: domain 1 of 1	20..145	68/129 (53%) 107/129 (83%)	8e-58

Example 36.

5　　　The NOV36 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 36A.

Table 36A. NOV36 Sequence Analysis		
	SEQ ID NO: 117	712 bp
NOV36a, CG58662-01 DNA Sequence	GCAGCTATTGCACTTAATCGCGGCTGCTAGCACCATGTCCCGCGTTTTGGTGCCTTGC CATGTGAAAGGCACCGTAGCCCTGCAGGTGGGGGACGTATGGACCTCCCAAGGCCGGC CTAGTGTGCTGGTCATTGATGTCACCTTCCCCTGTGTCACTCCGTTCGAGGGGATCAC ATTTAAGAATTATTACACAGCGTTTTTTGAGCATCCTGTCTGTCAGCACACCTCAGCA CACACACCGGCCAAGTGGGTGACCTGCCTGTGGGACTACTGTCTGATGCCCGACCCAC ACAGTGAGGAGGGAGCCCAGGAGTATGTGTCGCTGTTCAAGCAACAGATACTGTGTGA CATGGCCAGAATATCGGAGCTACACCTGATTCTGCAGCAGCCATCACCACTGTGGCTG TCTTTCACAGTGGAGGAGCTGCAGATCTATCAGCAGGGACCAAAGAGCCCCTCCATGA TCTTCCCCAAGTGGCTCTCCCACCCAGTGCCCTGTGAGCAACCTGCACTCCTCCATGA GGGTCTCCCAGACCCCAGCAGGGTATCCTCTGAGGTGCAGCAGATGTGGGCACTGACA GAGATGATCCGGGCCAGTCACACCTCCGCGAGGATAGGCCACTTTGATGTAGATGGCT GTTATGACCTGAACTTACTCTCCTACACTTGAGTGGTGGCTCCTAGCCAAGATGTTGG CCTTTCTGTGCCCACT
	ORF Start: ATG at 35	ORF Stop: TGA at 668
	SEQ ID NO: 118	211 aa	MW at 23932.3kD
NOV36a, CG58662-01 Protein Sequence	MSRVLVPCHVKGTVALQVGDVWTSQGRPSVLVIDVTFPCVTPFEGITFKNYYTAFFEH PVCQHTSAHTPAKWVTCLWDYCLMPDPHSEEGAQEYVSLFKQQILCDMARISELHLIL QQPSPLWLSFTVEELQIYQQGPKSPSMIFPKWLSHPVPCEQPALLHEGLPDPSRVSSE VQQMWALTEMIRASHTSARIGHFDVDGCYDLNLLSYT	
	SEQ ID NO: 119	843 bp
NOV36b, CG58662-02 DNA Sequence	CTGGCCTGAAGCCATGTCCCGCGTTCTAGCACCATGTCCCGCGTCTAGCACCATGTCC CGCGTCTAGCACCATGTCCCGCGTTCTAGCACCATGTCCCGCGTTCTAGCACCATGTC CCGCGTTCTAGCACCATGTCCCGCGTTTTGGTGCCTTGCCATGTGAAAGGCTCCGTAG CCCTCCAGGTGGGCGACGTGCGGACCTCCCAAGGCCGGCCTGGCGTGCTGGTCATCGA TGTCACCTTCCCCAGCGTCGCTCCCTTCGAGTTGCAGGAAATCACGTTTAAGAATTAC TACACAGCTTTTTTGAGCATCCGTGTCCGTCAGTACACCTCAGCACACACACCTGCCA AGTGGGTGACCTGCCTTCGGGACTACTGCCTGATGCCTGACCCACACAGTGAAGAGGG AGCCCAGGAGTATGTATCGCTGTTCAAGCATCAGATGCTATGTGACATGGCTAGAATA TCGGAGCTACGCCTGATTCTGCGGCAGCCATCACCACTGTGGCTGTCTTTCACAGTGG AGGAGCTGCAGATCTATCAGCAGGGACCAAAGAGCCCCTCCGTGACCTTTCCCAAGTG GCTCTCCCACCCAGTGCCCTGTGAGCAACCTGCACTCCTCCGTGAGGGTTTCCCAGAC CCCAGCAGGGTATCCTCCGAGGTGCAGCAGATGTGGGCACTGACAGAGATGATCCGGG CCAGTCACACCTCCGCAAGGATCGGCCGCTTTGATGTGGATGGCTGTTATGACCTGAC CTTGCTCTCCTACACTTGAATGGTTGCTCTTAGCCAAGATGTTGGCCTTTTTGTGGGC ACAGAAAGGCCAACGCGGGACATGGTGCTAG
	ORF Start: ATG at 132	ORF Stop: TGA at 771
	SEQ ID NO: 120	213 aa	MW at 24222.6kD
NOV36b, CG58662-02 Protein Sequence	MSRVLVPCHVKGSVALQVGDVRTSQGRPGVLVIDVTFPSVAPFELQEITFKNYYTAFL SIRVRQYTSAHTPAKWVTCLRDYCLMPDPHSEEGAQEYVSLFKHQMLCDMARISELRL ILRQPSPLWLSFTVEELQIYQQGPKSPSVTFPKWLSHPVPCEQPALLREGFPDPSRVS SEVQQMWALTEMIRASHTSARIGRFDVDGCYDLTLLSYT	

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 36B.

Table 36B. Comparison of NOV36a against NOV36b.		
Protein Sequence	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV36b	1..211 1..213	188/213 (88%) 193/213 (90%)

Further analysis of the NOV36a protein yielded the following properties shown in Table 36C.

209

Table 36C. Protein Sequence Properties NOV36a	
PSort analysis:	0.5666 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.1562 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV36a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 36D.

Table 36D. Geneseq Results for NOV36a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG04038	Human secreted protein, SEQ ID NO: 8119 - Homo sapiens, 115 aa. [EP1033401-A2, 06-SEP-2000]	1..103 1..105	82/105 (78%) 85/105 (80%)	1e-39

In a BLAST search of public sequence databases, the NOV36a protein was found to have homology to the proteins shown in the BLASTP data in Table 36E.

Table 36E. Public BLASTP Results for NOV36a				
Protein Accession Number	Protein/Organism/Length	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BSH3	SIMILAR TO RIKEN CDNA 1500032A17 GENE - Homo sapiens (Human), 213 aa.	1..211 1..213	190/213 (89%) 195/213 (91%)	e-107
Q9CQM0	1500032A17RIK PROTEIN - Mus musculus (Mouse), 213 aa.	1..211 1..213	174/213 (81%) 183/213 (85%)	4e-97

PFam analysis predicts that the NOV36a protein contains the domains shown in the Table 36F.

210

Table 36F. Domain Analysis of NOV36a			
Pfam Domain	NOV36a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 37.

The NOV37 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 37A.

Table 37A. NOV37 Sequence Analysis		
	SEQ ID NO: 121	520 bp
NOV37a, CG58584-01 DNA Sequence	CATTTGCTGTCTCCTCTGCTCACCAGCAGCTGTACTGGAGCCACCCGCGAAAATTCGG CCAGGGTTCTCGCTCTTGTCGTGTCTGTTCAAACCGGCACGGTCTGATCCGGAAATAT GGCCTCAATATGTGCCGCCAGTGTTTCCGTCAGTACGCGAAGGATATCGGTTTCATTA AGAAAGACCTGAGCTGTCTTCCTTGGCACTGCCTATGGAGGTGACACCCATCTCCTCC ATCATGGCCATCCTGAGACCGCTCGCGAAGCCCAAGATCATCAAAAAGAGCACCAAGT TCACTGGGAACCAGTCAGACTGATATGTCAAAATTAAGGGTAACTGGTGGAAACACAG AGGTATTGACAACAGGGTTCATAGAAGGTTTGAGGGCCAGATCTATGCCCAACATTGG TTATGGGAGAAACAAAAAGACAAAGCACATACTGCCCAGTGGCTTCTGGAAGTTCCTG GTCCACAACGTTAAGGAGCTGGAAGTACTGCTGGTGAGCAGAGGAGACAGCAAATG	
	ORF Start: TTT at 3	ORF Stop: TGA at 216
	SEQ ID NO: 122	71 aa	MW at 8461.8kD
NOV37a, CG58584-01 Protein Sequence	FAVSSAHQQLYWSHPRKFGQGSRSCRVCSNRHGLIRKYGLNMCRQCFRQYAKDIGFIK KDLSCLPWHCLWR	

Further analysis of the NOV37a protein yielded the following properties shown in
5   Table 37B.

Table 37B. Protein Sequence Properties NOV37a	
PSort analysis:	0.6400 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV37a protein against the Geneseq database, a proprietary databasê that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 37C.

Table 37C. Geneseq Results for NOV37a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG76128	Human colon cancer antigen protein SEQ ID NO:6892 - Homo sapiens, 80 aa. [WO200122920-A2, 05-APR-2001]	7..60 2..55	46/54 (85%) 48/54 (88%)	4e-24
AAM79084	Human protein SEQ ID NO 1746 - Homo sapiens, 56 aa. [WO200157190-A2, 09-AUG-2001]	7..60 3..56	39/54 (72%) 43/54 (79%)	2e-18
AAG39921	Arabidopsis thaliana protein fragment SEQ ID NO: 49464 - Arabidopsis thaliana, 637 aa. [EP1033405-A2, 06-SEP-2000]	7..63 3..58	40/57 (70%) 45/57 (78%)	2e-18
AAM80068	Human protein SEQ ID NO 3714 - Homo sapiens, 74 aa. [WO200157190-A2, 09-AUG-2001]	7..58 22..73	38/52 (73%) 42/52 (80%)	5e-18
AAG34802	Arabidopsis thaliana protein fragment SEQ ID NO: 42406 - Arabidopsis thaliana, 56 aa. [EP1033405-A2, 06-SEP-2000]	7..58 3..54	37/52 (71%) 42/52 (80%)	1e-17

In a BLAST search of public sequence databases, the NOV37a protein was found to have homology to the proteins shown in the BLASTP data in Table 37D.

212

Table 37D. Public BLASTP Results for NOV37a				
Protein Accession Number	Protein/Organism/Length	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
BAB79485	RIBOSOMAL PROTEIN S29 - Homo sapiens (Human), 56 aa.	7..60 3..56	53/54 (98%) 53/54 (98%)	1e-27
P30054	40S ribosomal protein S29 - Homo sapiens (Human),, 55 aa.	7..60 2..55	53/54 (98%) 53/54 (98%)	1e-27
Q90YP2	40S RIBOSOMAL PROTEIN S29 - Ictalurus punctatus (Channel catfish), 56 aa.	7..60 3..56	52/54 (96%) 53/54 (97%)	2e-27
AAL62474	RIBOSOMAL PROTEIN S29 - Spodoptera frugiperda (Fall armyworm), 56 aa.	7..60 3..56	41/54 (75%) 48/54 (87%)	6e-21
Q9VH69	CG8495 PROTEIN - Drosophila melanogaster (Fruit fly), 56 aa.	10..60 6..56	41/51 (80%) 46/51 (89%)	3e-20

PFam analysis predicts that the NOV37a protein contains the domains shown in the Table 37E.

Table 37E. Domain Analysis of NOV37a			
Pfam Domain	NOV37a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Ribosomal_S14: domain 1 of 1	7..61	17/60 (28%) 51/60 (85%)	7.5e-20

Example 38.

The NOV38 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 38A.

Table 38A. NOV38 Sequence Analysis		
	SEQ ID NO: 123	2039 bp
NOV38a, CG58538-01 DNA Sequence	GCAGCACACCTGCTCTGTGACTGACACTCTTGCAGAAGTGGGGCCACTTCAGGGACAT GGACAAGGTGTTGTACCTGCTGTCACAGAGCCTGTTATCTGTTCAGAATGACCGAAGA AGCATGCCGAACACGGAGTCAGAAACGAGCGCTTGAACGGGACCCAACAGAGGACGAT GTGGAGAGCAAGAAAATAAAAATGGAGAGAGGATTGTTGGCTTCAGATTTAAACACTG ACGGAGACATGAGGGTGACACCTGAGCCGGGAGCAGGTCCAACCCAAGGATTGCTGAG GGCAACAGAGGCCACGGCCATGGCCATGGGCAGAGGCGAAGGGCTGGTGGGCGATGGG CCCGTGGACATGCGCACCTCACACAGTGACATGAAGTCCGAGAGGAGACCCCCCTCAC CTGACGTGATTGTGCTCTCCGACAACGAGCAGCCCTCGAGCCCGAGAGTGAATGGGCT GACCACGGTGGCCTTGAAGGAGACTAGCACCGAGGCCCTCATGAAAAGCAGTCCTGAA	

213

	GAACGAGAAAGGATGATCAAGCAGCTGAAGGAAGAATTGAGGTTAGAAGAAGCAAAAC TCGTGTTGTTGAAAAAGTTGCGGCAGAGTCAAATACAAAAGGAAGCCACCGCCCAGAA GCCCACAGGTTCTGTTGGGAGCACCGTGACCACCCCTCCCCCGCTTGTTCGGGGCACT CAGAACATTCCTGCTGGCAAGCCATCACTCCAGACCTCTTCAGCTCGGATGCCCGGCA GTGTCATACCCCCGCCCCTGGTCCGAGGTGGGCAGCAGGCGTCCTCGAAGCTGGGGCC ACAGGCGAGCTCACAGGTCGTCATGCCCCCACTCGTCAGGGGGGCTCAGCAAATCCAC AGCATTAGGCAACATTCCAGCACAGGGCCACCGCCCCTCCTCCTGGCCCCCCGGGCGT CGGTGCCCAGTGTGCAGATTCAGGGACAGAGGATCATCCAGCAGGGCCTCATCCGCGT CGCCAATGTTCCCAACACCAGCCTGCTCGTCAACATCCCACAGCCCACCCCAGCATCA CTGAAGGGGACAACAGCCACCTCCGCTCAGGCCAACTCCACCCCCACTAGTGTGGCCT CTGTGGTCACCTCTGCCGAGTCTCCAGCAAGCCGACAGGCGGCCGCCAAGCTGGCGCT GCGCAAACAGCTGGAGAAGACGCTACTCGAGATCCCCCCACCCAAGCCCCCAGCCCCA GAGATGAACTTCCTGCCCAGCGCCGCCAACAACGAGTTCATCTACCTGGTCGGCCTGG AGGAGGTGGTGCAGAACCTACTGGAGACACAAGCAGGCAGGATGTCGGCCGCCACTGT GCTGTCCCGGGAGCCCTACATGTGTGCACAGTGCAAGACGGACTTCACGTGCCGCTGG CGGGAGGAGAAGAGCGGCGCCATCATGTGTGAGAACTGCATGACAACCAACCAGAAGA AGGCGCTCAAGGTGGAGCACACCAGCCGGCTGAAGGCCGCCTTTGTGAAGGCGCTGCA GCAGGAACAGGAGATTGAGCAGCGGCTCCTGCAGCAGGGCACGGCCCCTGCACAGGCC AAGGCCGAGCCCACCGCTGCCCCACACCCCGTGCTGAAGCAGGCCTCCAGCCAGCTGT CCCGGGGTTCGGCCACGACGCCCCGAGGTGTCCTGCACACGTTCAGTCCGTCACCCAA ACTGCAGAACTCAGCCTCGGCCACAGCCCTGGTCAGCAGGACCGGCAGACATTCTGAG AGAACCGTGAGCGCCGGCAAGGGCAGCGCCACCTCCAACTGGAAGAAGACGCCCCTCA GCACAGGCGGGACCCTTGCGTTTGTCAGCCCAAGCCTGGCGGTGCACAAGAGCTCCTC GGCCGTGGACCGCCAGCGAGAGTACCTCCTGGACATGATCCCACCCCGCTCCATCCCC CAGTCAGCCACGTGGAAATAGTGCGAGCCAGGCCCCGTGGAAGACGGGCTCCCTCCTC CCCCACCTGGCCCCTGGTCTAGAAGGACCCACTGCACCACCCTCCGCTGGCTCGGGAA GACACCGTG
ORF Start: ATG at 106	**ORF Stop: TAG at 1933**
SEQ ID NO: 124	**609 aa**	**MW at 65295.8kD**
NOV38a, CG58538-01 Protein Sequence	MTEEACRTRSQKRALERDPTEDDVESKKIKMERGLLASDLNTDGDMRVTPEPGAGPTQ GLLRATEATAMAMGRGEGLVGDGPVDMRTSHSDMKSERRPPSPDVIVLSDNEQPSSPR VNGLTTVALKETSTEALMKSSPEERERMIKQLKEELRLEEAKLVLLKKLRQSQIQKEA TAQKPTGSVGSTVTTPPPLVRGTQNIPAGKPSLQTSSARMPGSVIPPPLVRGGQQASS KLGPQASSQVVMPPLVRGAQQIHSIRQHSSTGPPPLLLAPRASVPSVQIQGQRIIQQG LIRVANVPNTSLLVNIPQPTPASLKGTTATSAQANSTPTSVASVVTSAESPASRQAAA KLALRKQLEKTLLEIPPPKPPAPEMNFLPSAANNEFIYLVGLEEVVQNLLETQAGRMS AATVLSREPYMCAQCKTDFTCRWREEKSGAIMCENCMTTNQKKALKVEHTSRLKAAFV KALQQEQEIEQRLLQQGTAPAQAKAEPTAAPHPVLKQASSQLSRGSATTPRGVLHTFS PSPKLQNSASATALVSRTGRHSERTVSAGKGSATSNWKKTPLSTGGTLAFVSPSLAVH KSSSAVDRQREYLLDMIPPRSIPQSATWK

Further analysis of the NOV38a protein yielded the following properties shown in Table 38B.

Table 38B. Protein Sequence Properties NOV38a	
PSort analysis:	0.4404 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1257 probability located in mitochondrial inner membrane; 0.1257 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV38a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 38C.

214

Table 38C. Geneseq Results for NOV38a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV38a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAM00991	Human bone marrow protein, SEQ ID NO: 492 - Homo sapiens, 502 aa. [WO200153453-A2, 26-JUL-2001]	1..471 4..473	217/504 (43%) 290/504 (57%)	2e-87
AAM00944	Human bone marrow protein, SEQ ID NO: 420 - Homo sapiens, 546 aa. [WO200153453-A2, 26-JUL-2001]	1..471 48..517	217/504 (43%) 290/504 (57%)	2e-87
AAM00831	Human bone marrow protein, SEQ ID NO: 194 - Homo sapiens, 266 aa. [WO200153453-A2, 26-JUL-2001]	1..197 47..262	84/217 (38%) 110/217 (49%)	1e-23
AAM85818	Human immune/haematopoietic antigen SEQ ID NO:13411 - Homo sapiens, 84 aa. [WO200157182-A2, 09-AUG-2001]	417..471 1..55	41/55 (74%) 49/55 (88%)	7e-19

In a BLAST search of public sequence databases, the NOV38a protein was found to have homology to the proteins shown in the BLASTP data in Table 38D.

Table 38D. Public BLASTP Results for NOV38a				
Protein Accession Number	**Protein/Organism/Length**	**NOV38a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
No Significant Matches Found				

PFam analysis predicts that the NOV38a protein contains the domains shown in the Table 38E.

Table 38E. Domain Analysis of NOV38a			
Pfam Domain	**NOV38a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
GATA: domain 1 of 1	414..453	12/43 (28%) 17/43 (40%)	1.1

215

Example 39.

The NOV39 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 39A.

Table 39A. NOV39 Sequence Analysis		
	SEQ ID NO: 125	1421 bp
NOV39a, CG59371-01 DNA Sequence	ACCATTTCAGAGATGTCTTCCAGAAGTACCAAAGATTTAATTAAAAGTAAGTGGGGAT CGAAGCCTAGTAACTCCAAATCCGAAACTACATTAGAAAAATTAAAGGGAGAAATTGC ACACTTAAAGACATCAGTGGATGAAATCACAAGTGGGAAAGGAAAGCTGACTGATAAA GAGAGACACAGACTTTTGGAGAAAATTCGAGTCCTTGAGGCTGAGAAGGAGAAGAATG CTTATCAACTCACAGAGAAGGACAAAGAAATACAGCGACTGAGAGACCAACTGAAGGC CAGATATAGTACTACCACATTGCTTGAACAGCTGGAAGAGACAACGAGAGAAGGAGAA AGGAGGGAGCAGGTGTTGAAAGCCTTATCTGAAGAGAAAGACGTATTGAAACAACAGT TGTCTGCTGCAACCTCACGAATTGCTGAACTTGAAAGCAAAACCAATACACTCCGTTT ATCACAGACTGTGGCTCCAAACTGCTTCAACTCATCAATAAATAATATTCATGAAATG GAAATACAGCTGAAAGATGCTCTGGAGAAAAATCAGCAGTGGCTCGTGTATGATCAGC AGCGGGAAGTCTATGTAAAAGGACTTTTAGCAAAGATCTTTGAGTTGGAAAAGAAAAC GGAAACAGCTGCTCATTCACTCCCACAGCAGACAAAAAAGCCTGAATCAGAAGGTTAT CTTCAAGAAGAGAAGCAGAAATGTTACAACGATCTCTTGGCAAGTGCAAAAAAAGATC TTGAGGTTGAACGACAAACCATAACTCAGCTGAGTTTTGAACTGAGTGAATTTCGAAG AAAATATGAAGAAACCCAAAAAGAAGTTCACAATTTAAATCAGCTGTTGTATTCACAA AGAAGGGCAGATGTGCAACATCTGGAAGATGATAGGCATAAAACAGAGAAGATACAAA AACTCAGGGAAGAGAATGATATTGCTAGGGGAAAACTTGAAGAAGAGAAGAAGAGATC CGAAGAGCTCTTATCTCAGGTCCAGTCTCTTTACACATCTCTGCTAAAGCAGCAAGAA GAACAAACAAGGGTAGCTCTGTTGGAACAACAGATGCAGGCATGTACTTTAGACTTTG AAAATGAAAAACTCGACCGTCAACATGTGCAGCATCAATTGCATGTAATTCTTAAGGA GCTCCGAAAAGCAAGAAAAAATATAACACAGTTGGAATCCTTGAAACAGCTTCATGAG TTTGCCATCACAGAGCCATTAGTCACTTTCCAAGGAGAGACTGAAAACAGAGAAAAAG TTGCCGCCTCACCAAAAAGTCCCACTGCTGCACTCAATGGAAGCCTGGTGGAATGTCC CAAGTGCAATATACAGTATCCAGCCACTGAGCATCGCGATCTGCTTGTCCATGTGGAA TACTGTTCAAAGTAGCAAAATAAGTATTT

	ORF Start: ATG at 13	ORF Stop: TAG at 1405	
	SEQ ID NO: 126	464 aa	MW at 54045.6kD
NOV39a, CG59371-01 Protein Sequence	MSSRSTKDLIKSKWGSKPSNSKSETTLEKLKGEIAHLKTSVDEITSGKGKLTDKERHR LLEKIRVLEAEKEKNAYQLTEKDKEIQRLRDQLKARYSTTTLLEQLEETTREGERREQ VLKALSEEKDVLKQQLSAATSRIAELESKTNTLRLSQTVAPNCFNSSINNIHEMEIQL KDALEKNQQWLVYDQQREVYVKGLLAKIFELEKKTETAAHSLPQQTKKPESEGYLQEE KQKCYNDLLASAKKDLEVERQTITQLSFELSEFRRKYEETQKEVHNLNQLLYSQRRAD VQHLEDDRHKTEKIQKLREENDIARGKLEEEKKRSEELLSQVQSLYTSLLKQQEEQTR VALLEQQMQACTLDFENEKLDRQHVQHQLHVILKELRKARKNITQLESLKQLHEFAIT EPLVTFQGETENREKVAASPKSPTAALNGSLVECPKCNIQYPATEHRDLLVHVEYCSK		

Further analysis of the NOV39a protein yielded the following properties shown in Table 39B.

Table 39B. Protein Sequence Properties NOV39a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV39a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 39C.

Table 39C. Geneseq Results for NOV39a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB92925	Human protein sequence SEQ ID NO:11576 - Homo sapiens, 231 aa. [EP1074617-A2, 07-FEB-2001]	170..392 1..223	222/223 (99%) 222/223 (99%)	e-122
AAG75490	Human colon cancer antigen protein SEQ ID NO:6254 - Homo sapiens, 165 aa. [WO200122920-A2, 05-APR-2001]	1..67 99..165	64/67 (95%) 64/67 (95%)	1e-28
AAM78520	Human protein SEQ ID NO 1182 - Homo sapiens, 990 aa. [WO200157190-A2, 09-AUG-2001]	6..394 515..929	96/421 (22%) 182/421 (42%)	3e-12
AAM41000	Human polypeptide SEQ ID NO 5931 - Homo sapiens, 1988 aa. [WO200153312-A1, 26-JUL-2001]	70..420 852..1203	90/384 (23%) 161/384 (41%)	3e-12
AAM40999	Human polypeptide SEQ ID NO 5930 - Homo sapiens, 1988 aa. [WO200153312-A1, 26-JUL-2001]	70..420 852..1203	90/384 (23%) 161/384 (41%)	3e-12

In a BLAST search of public sequence databases, the NOV39a protein was found to
5   have homology to the proteins shown in the BLASTP data in Table 39D.

Table 39D. Public BLASTP Results for NOV39a				
Protein Accession Number	Protein/Organism/Length	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96H32	SIMILAR TO RIKEN CDNA 1200008O12 GENE - Homo sapiens (Human), 464 aa.	1..464 1..464	458/464 (98%) 458/464 (98%)	0.0
Q9DBZ8	1200008O12RIK PROTEIN - Mus musculus (Mouse), 462 aa.	1..464 1..462	348/464 (75%) 401/464 (86%)	0.0

217

Q9NVS7	CDNA FLJ10540 FIS, CLONE NT2RP2001245 - Homo sapiens (Human), 231 aa.	170..392 1..223	222/223 (99%) 222/223 (99%)	e-122
Q9CZP8	2700032M20RIK PROTEIN - Mus musculus (Mouse), 189 aa.	1..176 1..176	121/176 (68%) 150/176 (84%)	3e-63
Q9VJE5	CLIP-190 PROTEIN - Drosophila melanogaster (Fruit fly), 1690 aa.	4..439 675..1118	108/461 (23%) 203/461 (43%)	2e-16

PFam analysis predicts that the NOV39a protein contains the domains shown in the Table 39E.

Table 39E. Domain Analysis of NOV39a			
Pfam Domain	NOV39a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 40.

The NOV40 clone was analyzed, and the nucleotide and predicted polypeptide
sequences are shown in Table 40A.

Table 40A. NOV40 Sequence Analysis	
	SEQ ID NO: 127 3955 bp
NOV40a, CG59346-01 DNA Sequence	TGCACCCTCGCTGCCTCCTTTCCTCCATGCTGCCTGGATCTGGCGAGCTGGGGTGATT AATTGGCTATGATGATGAACGTCCCCGGCGGAGGAGCGGCCGCGGTGATGATGACGGG CTACAATAATGGTCGCTGTCCCCGGAATTCTCTCTACAGTGACTGCATTATTGAGGAG AAGACGGTGGTCCTGCAGAAAAAAGACAATGAGGGCTTTGGATTCGTGCTTCGAGGGG CCAAAGCTGACACACCCATTGAAGAATTCACACCAACACCGGCTTTCCCAGCCCTACA GTACCTGGAGTCCGTGGATGAAGGTGGGGTGGCGTGGCAAGCCGGACTAAGGACCGGG GACTTCTTGATTGAGGTTAACAATGAGAATGTTGTCAAAGTCGGCCACAGGCAGGTGG TGAACATGATCCGGCAGGGAGGGAATCACCTGGTCCTTAAGGTGGTCACGGTGACCAG GAATCTGGACCCCGACGACACCGCCAGGAAGAAAGCTCCCCCGCCTCCAAAGCGGGCA CCGACCACAGCCCTCACCCTGCGCTCCAAGTCCATGACCTCGGAGCTGGAGGAGCTCG ATAAACCCGAGGAGATAGTCCCGGCCTCCAAGCCCTCCCGCGCTGCTGAGAACATGGC TGTGGAACCGAGGGTGGCGACCATCAAGCAGCGGCCCAGCAGCCGGTGCTTCCCGGCG GGCTCAGACATGAACGTGAGTGGCCGTACCTTGGGACCACGAGGGCGGGGGCCGACGG TGCCCCCTAGGCTCTCTGGTTTGCAGTCTGTGTACGAACGCCAAGGAATCGCCGTGAT GACGCCCACTGTTCCTGGGAGCCCAAAAGCCCCGTTTCTGGGCATCCCTCGAGGTACG ATGCGAAGGCAGAAATCAATAGGAATAACAGAGGAAGAGCGGCAGTTTCTGGCTCCTC CAATGCTGAAGTTCACCAGAAGCCTGTCCATGCCGGACACCTCTGAGGACATCCCCCC TCCACCGCAGTCTGTGCCCCCGTCCCCACCACCACCTTCCCCAACCACTTACAACTGC CCCAAGTCCCCAACTCCAAGAGTCTACGGGACGATTAAGCCTGCGTTCAATCAGAATT CTGCCGCCAAGGTGTCCCCCGCCACCAGGTCCGACACCGTGGCCACCATGATGAGGGA GAAGGGGATGTACTTCAGGAGAGAGCTGGACCGCTACTCCTTGGACTCTGAAGACCTC TACAGTCGGAATGCCGGCCCGCAAGCCAACTTCCGCAACAAGAGAGGCCAGATGCCAG AAAACCCATACTCAGAGGTGGGGAAGATCGCCAGCAAAGCCGTCTACGTCCCCGCCAA GCCCGCCAGGCGGAAGGGGATGCTGGTGAAGCAGTCCAACGTGGAGGACAGCCCCGAG AAGACGTGCTCCATCCCTATCCCGACCATCATCGTGAAGGAGCCGTCCACCAGCAGCA GCGGCAAGAGCAGCCAGGGCAGCAGCATGGAGATCGACCCCCAGGCCCCGGAGCCACC GAGCCAGCTGCGGCCTGACGAAAGCCTGACCGTCAGCAGCCCCTTTGCCGCCGCCATC GCCGGAGCCGTCCGCGACCGTGAGAAGCGGCTGGAAGCCAGGAGGAACTCCCCGGCCT TCCTCTCCACAGACCTGGGGGATGAGGATGTGGGCCTGGGGCCACCCGCCCCCAGGAC GCGGCCCTCCATGTTCCCCGAGGAGGGGGATTTTGCTGACGAGGACAGCGCTGAGCAG CTGTCATCCCCCATGCCGAGTGCCACGCCCAGGGAGCCCGAAAACCATTTCGTGGGTG GCGCCGAGGCCAGTGCTCCGGGTGAGGCTGGGAGGCCGCTGAATTCCACGTCCAAAGC

218

	CCAGGGGCCCGAGAGCAGCCCAGCAGTGCCCTCCGCGAGCAGCGGCACAGCCGGCCCC GGGAATTATGTCCACCCACTCACAGGGCGGCTGCTTGATCCCAGCTCCCCGCTGGCCC TGGCACTCTCCGCAAGGGACCGAGCCATGAAGGAGTCTCAACAGGGACCCAAAGGGGA GGCCCCCAAGGCCGACCTCAACAAACCTCTTTACATTGATACCAAAATGCGGCCCAGC CTGGATGCCGGCTTCCCTACGGTCACCAGGCAGAACACCCGGGGACCCCTGAGGCGGC AGGAGACGGAGAACAAGTACGAGACCGACCTGGGCCGAGACCGGAAAGGCGATGACAA GAAGAACATGCTGATCGACATCATGGACACGTCCCAGCAGAAGTCGGCTGGCCTGCTG ATGGTGCACACCGTGGACGCCACTAAGCTGGACAACGCCCTGCAGGAAGAGGACGAGA AGGCAGAGGTGGAGATGAAGCCAGACAGCTCGCCGTCCGAGGTGCCAGAAGGTGTTTC CGAAACCGAAGGTGCTTTACAGATCTCCGCTGCCCCCGAGCCCACCACCGTGCCCGGC AGAACCATCGTCGCGGTGGGCTCCATGGAAGAGGCGGTGATTTTGCCATTCCGCATCC CTCCTCCCCCTCTGGCATCCGTGGACTTGGATGAGGATTTTATTTTTACAGAGCCATT GCCTCCTCCCCTGGAATTTGCAAATAGTTTTGATATCCCCGATGACCGGGCAGCTTCT GTCCCGGCTCTCTCAGACTTAGTGAAGCAGAAGAAAAGCGACACCCCTCAGTCCCCTT CGTTGAACTCCAGCCAACCAACCAACTCTGCAGCAGCAAGAAGCCAGCCAGTCTTTC AAACTGTCTGCCTGCCTCATTCCTGCCACCCCCTGAAAGCTTTGACGCCGTCGCCGAC TCTGGGATCGAGGAGGTGGACAGCCGGAGTAGCAGCGACCACCACCTCGAGACGACCA GCACTATCTCCACCGTGTCTAGCATCTCCACCCTGTCTTCCGAAGGTGGAGAGAATGT GGACACCTGCACAGTCTATGCAGATGGGCAAGCATTTATGGTTGACAAACCCCCAGTA CCTCCTAAGCCAAAAATGAAGCCCATCATTCACAAAAGCAATGCACTTTATCAAGACG CGCTCGTGGAAGAAGATGTAGATAGCTTTGTTATCCCCCCGCCCGCTCCCCCGCCCCC GCCGGGCAGTGCCCAGCCTGGGATGGCCAAGGTTCTCCAGCCAAGGACCTCCAAGTTG TGGGGCGACGTCACAGAGATCAAAAGCCCGATTCTCTCAGGCCCAAAGGCAAACGTTA TTAGTGAATTGAACTCTATCCTACAGCAAATGAACCGAGAGAAATTGGCAAAGCCGGG GGAAGGACTGGATTCACCAATGGGAGCCAAGTCCGCCAGCCTCGCTCCAAGAAGCCCG GAGATCATGAGCACCATCTCAGGTACACGGAGCACGACGGTCACCTTCACTGTTCGCC CCGGCACCTCCCAGCCCATCACCCTGCAGAGCCGGCCCCCCGACTATGAAAGCAGGAC CTCAGGAACAAGACGTGCCCCAAGCCCTGTGGTCTCGCCAACAGAGATGAACAAAGAG ACCCTGCCCGCCCCCTGTCTGCTGCCACCGCCTCTCCTTCTCCCGCTCTCTCAGATG TCTTTAGCCTTCCAAGCCAGCCCCCTTCTGGGGATCTATTTGGCTTGAACCCAGCGGG ACGCAGTAGGTCGCCATCCCCCTCGATACTGCAACAGCCAATCTCAAATAAGCCTTTT ACAACTAAACCTGTCCACCTGTGGACTAAACCAGATGTGGCCGATTGGCTGGAAAGTC TAAACTTGGGTGAACATAAAGAGGCCTTCATGGCAATGAGATCGATGGCAGTCACTT ACCAAACCTGCAGAAGGAGGACCTCATCGATCTTGGGGTAACTCGAGTCGGGCACAGA ATGAACATAGAAAGGGCTTTGAAACAGCTGCTGGACAGATAAGGACGGCTGCTCTCCA CCTCGCAGACTGCTCTTGTTATAAGTAGAGATGGGCTCGTGCTGAAACATCTGAATGC CAAGCGAAGTC
	ORF Start: ATG at 67	ORF Stop: TAA at 3868
	SEQ ID NO: 128	1267 aa	MW at 136108.7kD
NOV40a, CG59346-01 Protein Sequence	MMMNVPGGGAAAVMMTGYNNGRCPRNSLYSDCIIEEKTVVLQKKDNEGFGFVLRGAKA DTPIEEFTPTPAFPALQYLESVDEGGVAWQAGLRTGDFLIEVNNENVVKVGHRQVVNM IRQGGNHLVLKVVTVTRNLDPDDTARKKAPPPPKRAPTTALTLRSKSMTSELEELDKP EEIVPASKPSRAAENMAVEPRVATIKQRPSSRCFPAGSDMNVSGRTLGPRGRGPTVPP RLSGLQSVYERQGIAVMTPTVPGSPKAPFLGIPRGTMRRQKSIGITEEERQFLAPPML KFTRSLSMPDTSEDIPPPPQSVPPSPPPPSPTTYNCPKSPTPRVYGTIKPAFNQNSAA KVSPATRSDTVATMMREKGMYFRRELDRYSLDSEDLYSRNAGPQANFRNKRGQMPENP YSEVGKIASKAVYVPAKPARRKGMLVKQSNVEDSPEKTCSIPIPTIIVKEPSTSSSGK SSQGSSMEIDPQAPEPPSQLRPDESLTVSSPFAAAIAGAVRDREKRLEARRNSPAFLS TDLGDEDVGLGPPAPRTRPSMFPEEGDFADEDSAEQLSSPMPSATPREPENHFVGGAE ASAPGEAGRPLNSTSKAQGPESSPAVPSASSGTAGPGNYVHPLTGRLLDPSSPLALAL SARDRAMKESQQGPKGEAPKADLNKPLYIDTKMRPSLDAGFPTVTRQNTRGPLRRQET ENKYETDLGRDRKGDDKKNMLIDIMDTSQQKSAGLLMVHTVDATKLDNALQEEDEKAE VEMKPDSSPSEVPEGVSETEGALQISAAPEPTTVPGRTIVAVGSMEEAVILPFRIPPP PLASVDLDEDFIFTEPLPPPLEFANSFDIPDDRAASVPALSDLVKQKKSDTPQSPSLN SSQPTNSADSKKPASLSNCLPASFLPPPESFDAVADSGIEEVDSRSSSDHHLETTSTI STVSSISTLSSEGGENVDTCTVYADGQAFMVDKPPVPPKPKMKPIIHKSNALYQDALV EEDVDSFVIPPPAPPPPPGSAQPGMAKVLQPRTSKLWGDVTEIKSPILSGPKANVISE LNSILQQMNREKLAKPGEGLDSPMGAKSASLAPRSPEIMSTISGTRSTTVTFTVRPGT SQPITLQSRPPDYESRTSGTRRAPSPVVSPTEMNKETLPAPLSAATASPSPALSDVFS LPSQPPSGDLFGLNPAGRSRSPSPSILQQPISNKPFTTKPVHLWTKPDVADWLESLNL GEHKEAFMDNEIDGSHLPNLQKEDLIDLGVTRVGHRMNIERALKQLLDR

Further analysis of the NOV40a protein yielded the following properties shown in Table 40B.
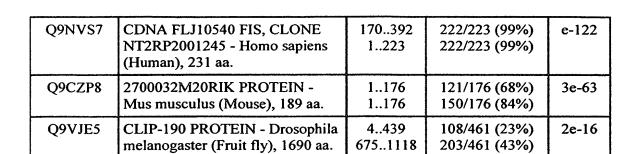
Table 40B. Protein Sequence Properties NOV40a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV40a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 40C.

Table 40C. Geneseq Results for NOV40a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM79240	Human protein SEQ ID NO 1902 - Homo sapiens, 1248 aa. [WO200157190-A2, 09-AUG-2001]	14..1267 1..1248	1231/1271 (96%) 1231/1271 (96%)	0.0
AAB31518	Amino acid sequence of the rat Shank2 polypeptide - Rattus sp, 1470 aa. [WO200078921-A2, 28-DEC-2000]	30..1267 240..1470	1078/1255 (85%) 1132/1255 (89%)	0.0
AAM80224	Human protein SEQ ID NO 3870 - Homo sapiens, 1161 aa. [WO200157190-A2, 09-AUG-2001]	172..1267 82..1161	1071/1103 (97%) 1071/1103 (97%)	0.0
AAB31517	Amino acid sequence of the rat Shank3a polypeptide - Rattus sp, 1740 aa. [WO200078921-A2, 28-DEC-2000]	18..1264 550..1737	496/1349 (36%) 673/1349 (49%)	0.0
AAY83017	Rat shank 3a - Rattus rattus, 1740 aa. [WO200011204-A2, 02-MAR-2000]	18..1264 550..1737	496/1349 (36%) 673/1349 (49%)	0.0

In a BLAST search of public sequence databases, the NOV40a protein was found to have homology to the proteins shown in the BLASTP data in Table 40D.

Table 40D. Public BLASTP Results for NOV40a				
Protein Accession Number	Protein/Organism/Length	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UPX8	KIAA1022 PROTEIN - Homo sapiens (Human), 1131 aa (fragment).	124..1267 1..1131	1121/1154 (97%) 1121/1154 (97%)	0.0
Q9QX93	PROLINE RICH SYNAPSE ASSOCIATED PROTEIN 1 - Rattus norvegicus (Rat), 1252 aa.	2..1267 1..1252	1103/1276 (86%) 1158/1276 (90%)	0.0
O70470	CORTACTIN-BINDING PROTEIN 1 - Rattus norvegicus (Rat), 1252 aa.	2..1267 1..1252	1102/1276 (86%) 1158/1276 (90%)	0.0
Q9WUV9	PROLINE RICH SYNAPSE ASSOCIATED PROTEIN 1 - Rattus norvegicus (Rat), 1259 aa.	2..1267 1..1259	1103/1283 (85%) 1158/1283 (89%)	0.0
Q9WUW0	PROLINE RICH SYNAPSE ASSOCIATED PROTEIN 1 - Rattus norvegicus (Rat), 1250 aa.	2..1267 1..1250	1095/1276 (85%) 1151/1276 (89%)	0.0

PFam analysis predicts that the NOV40a protein contains the domains shown in the Table 40E.

Table 40E. Domain Analysis of NOV40a			
Pfam Domain	NOV40a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PDZ: domain 1 of 1	38..131	23/97 (24%) 70/97 (72%)	1e-07
SAM: domain 1 of 1	1202..1265	27/68 (40%) 53/68 (78%)	9.8e-22

Example 41.

5   The NOV41 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 41A.

Table 41A. NOV41 Sequence Analysis		
	SEQ ID NO: 129	2069 bp
NOV41a, CG57814-01 DNA Sequence	GGACACTGACATGGACTGAAGGAGTAGAAAGCACTATAAATGTCTTTCCTTATCTGTG TGTACTCTTATCTCACTGTTCTATTTTTTCTCCTCATTTATATTAACTCTTTCTTACC TTTTTTTCTGAACTTCTAGGCCTTCTCTTTCCAGAACTGGTGGAAGACAAATGAAACG GCCAAGATGGTAAGAAACAAGCCGCATTTCTCCTTGGGGAGACTGATAATTTAAAAGG TTTGTTGTGTCAGAAACATTCCCAGCTTCATCACCAACCCTTTCCTTCCACCTCTGCC CACTGGAGACCACTTACATCCCGAAGCGGACGCGGCAGCTGAAGTCAGGAAACCATGC ATCACATTAGCAGGAGCCAACTGCAGACTTTAAACTCCGTTCAACATGTGGATGCGGC AGAGAAATGACCTGTCCAGACAAGCCGGGGCAGCTCATAAACTGGTTCATCTGCTCCC TGTGCGTCCCGCGGGTGCGTAAGCTCTGGAGCAGCCGGCGTCAAGGACCCGGAGAAA CCTTCTGCTGGGCACTGCGTGTGCCATCTACTTGGGCTTCCTGGTGAGCCAGGTGGGG AGGGCCTCTCTCCAGCATGGACAGGCGGCTGAGAAGGGGCCCACATCGCAGCCGCGACA CCGCCGAGCCATCCTTCCCTGAGATACCCCTGGATGGTACCCTGGCCCCTCCAGAGTC CCAGGGCAATGGGTCCACTCTGCAGCCCAATGTGGTGTACATTACCCTACGCTCCGAG CGCAGCAAGCCGGCCAATATCCGTGGCACCGTGAAGCCCAAGCGCAGGAAAAAGCATG CAGTGGCATCGGCTGCCCCAGGGCAGGAGGCTTTGGTCGGACCATCCCTTCAGCCGCA GGAAGCGGCAAGGGAAGCTGATGCTGTAGCACCTGGGTACGCTCAGGGAGCAAACCTG GTTAAGATTGGAGAGCGACCCTGGAGGTTGGTGCGGGGTCCGGGAGTGCGAGCCGGGG GCCCAGACTTCCTGCAGCCCAGCTCCAGGGAGAGCAACATTAGGATCTACAGCGAGAG CGCCCCCTCCTGGCTGAGCAAAGATGACATCCGAAGAATGCGACTCTTGGCGGACAGC GCAGTGGCAGGGCTCCGGCCTGTGTCCTCTAGGAGCGGAGCCCGTTTGCTGGTGCTGG AGGGGGGCGCACCTGGCGCTGTGCTCCGCTGTGGCCCTAGCCCCTGTGGGCTTCTCAA GCAGCCCTTGGACATGAGTGAGGTGTTTGCCTTCCACCTAGACAGGATCCTGGGGCTC AACAGGACCCTGCCGTCTGTGAGCAGGAAAGCAGAGTTCATCCAAGATGGCCGCCCAT GCCCCATCATTCTTTGGGATGCATCTTTATCTTCAGCAAGTAATGACACCCATTCTTC TGTTAAGCTCACCTGGGGAACTTATCAGCAGTTGCTGAAACAGAAATGCTGGCAGAAT GGCCGAGTACCCAAGCCTGAATCAGGTTGTACTGAAATACATCATCATGAGTGGTCCA AGATGGCACTCTTTGATTTTTTGTTACAGATTTATAATCGCTTAGATACAAATTGCTG TGGATTCAGACCTCGCAAGGAAGATGCCTGTGTACAGAATGGATTGAGGCCAAAATGT GATGACCAAGGTTCTGCGGCTCTAGCACACATTATCCAGCGAAAGCATGACCCAAGGC ATTTGGTTTTTATAGACAACAAGGGTTTCTTTGACAGGAGTGAAGATAACTTAAACTT CAAATTGTTAGAAGGCATCAAAGAGTTTCCAGCTTCTGCAGTTTCTGTTTTGAAGAGC CAGCACTTACGGCAGAAACTTCTTCAGTCTCTGTTTCTTGATAAAGTGTATTGGGAAA GTCAAGGAGGTAGACAAGGAATTGAAAAGCTTATCGATGTAATAGAACACAGAGCCAA AATTCTTATCACCTATATCAATGCACACGGGGTCAAAGTATTACCTATGAATGAATGA CAAAAGAATCTTCTGGCTAGGGTGTTAGATATATTTATGCATTTTTGGTTTTGTTTTT AAATCAAGCACATCAACCTCAAGCCCGTTTAGCAATGAG
	ORF Start: ATG at 413	ORF Stop: TGA at 1970
	SEQ ID NO: 130	519 aa	MW at 57552.4kD
NOV41a, CG57814-01 Protein Sequence	MTCPDKPGQLINWFICSLCVPRVRKLWSSRRPRTRRNLLLGTACAIYLGFLVSQVGRA SLQHGQAAEKGPHRSRDTAEPSFPEIPLDGTLAPPESQGNGSTLQPNVVYITLRSERS KPANIRGTVKPKRRKKHAVASAAPGQEALVGPSLQPQEAAREADAVAPGYAQGANLVK IGERPWRLVRGPGVRAGGPDFLQPSSRESNIRIYSESAPSWLSKDDIRRMRLLADSAV AGLRPVSSRSGARLLVLEGGAPGAVLRCGPSPCGLLKQPLDMSEVFAFHLDRILGLNR TLPSVSRKAEFIQDGRPCPIILWDASLSSASNDTHSSVKLTWGTYQQLLKQKCWQNGR VPKPESGCTEIHHHEWSKMALFDFLLQIYNRLDTNCCGFRPRKEDACVQNGLRPKCDD QGSAALAHIIQRKHDPRHLVFIDNKGFFDRSEDNLNFKLLEGIKEFPASAVSVLKSQH LRQKLLQSLFLDKVYWESQGGRQGIEKLIDVIEHRAKILITYINAHGVKVLPMNE	
	SEQ ID NO: 131	1740 bp
NOV41b, CG57814-02 DNA Sequence	GGCAGCTGAAGTCAGGAAACCATGCATCACATTAGCAGGAGCCAACTGCAGACTTTAA ACTCCGTTCAACATGTGGATGCGGCAGAGAAATGACCTGTCCAGACAAGCCGGGGCAG CTCATAAACTGGTTCATCTGCTCCCTGTGCGTCCCGCGGGTGCGTAAGCTCTGGAGCA GCCGGCGTCCAAGGACCCGGAGAAACCTTCTGCTGGGCACTGCGTGTGCCATCTACTT GGGCTTCCTGGTGAGCCAGGTGGGGAGGGCCTCTCTCCAGCATGGACAGGCGGCTGAG AAGGGGCCACATCGCAGCCGCGACACCGCCGAGCCATCCTTCCCTGAGATACCCCTGG ATGGTACCCTGGCCCCTCCAGAGTCCCAGGGCAATGGGTCCACTCTGCAGCCCAATGT GGTGTACATTACCCTACGCTCCAAGCGCAGCAAGCCGGCCAATATCCGTGGCACCGTG AAGCCCAAGCGCAGGAAAAAGCATGCAGTGGCATCGGCTGCCCAAGGGCAGGAGGCTT TGGTCGGACCATCCCTTCAGCCGCAAGAAGCGGCAAGGGAAGCTGATGCTGTAGCACT GGGTACGCTCAGGAGCAAACTGGTTAAGATGGAGAGCGACCCTGAAGGTGGTGCGGGG TCGGGAGTGCGAGCCGGGGGCCCAGACTTCCTGCAGCCCAGCTCCAGGGAGAGCAACA TTAGGATCTACAGCGAGAGCGCCCCCTCCTGGCTGAGCAAAGATGACATCCGAAGAAT GCGACTCTTGGCGGACAGCGCAGTGGCTCCGGCCTCCGGCCTGTGTCCTCTAGGAGCGGA GCCCGTTTGCTGGTGCTGGAGGGGGGCGCACCTGGCGCTGTGCTCCGCTGTGGCCCTA GCCCCTGTGGGCTTCTCAAGCAGCCCTTGGACATGAGTGAGGTGTTTGCCTTCCACCT AGACAGGATCCTGGGGCTCAACAGGACCCTGCCGTCTGTGAGCAGGAAAGCAGAGTTC

222

	ATCCAAGATGGCCGCCCATGCCCCATCATTCTTTGGGATGCATCTTTATCTTCAGCAA GTAATGACACCCATTCTTCTGTTAAGCTCACCTGGGGAACTTATCAGCAGTTGCTGAA ACAGAAATGCTGGCAGAATGGCCGAGTACCCAAGCCTGAATCAGGTTGTACTGAAATA CATCATCATGAGTGGTCCAAGATGGCACTCTTTGATTTTTTGTTACAGATTTATAATC GCTTAGATACAAATTGCTGTGGATTCAGACCTCGCAAGGAAGATGCCTGTGTACAGAA TGGATTGAGGCCAAAATGTGATGACCAAGGTTCTGCGGCTCTAGCACACATTATCCAG CGAAAGCATGACCCAAGGCATTTGGTTTTTATAGACAACAAGGGTTTCTTTGACAGGA GTGAAGATAACTTAAACTTCAAATTGTTAGAAGGCATCAAAGAGTTTCCAGCTTCTGC AGTTTCTGTTTTGAAGAGCCAGCACTTACGGCAGAAACTTCTTCAGTCTCTGTTTCTT GATAAAGTGTATTGGGAAAGTCAAGGAGGTAGACAAGGAATTGAAAAGCTTATCGATG TAATAGAACACAGAGCCAAAATTCTTATCACCTATATCAATGCACACGGGGTCAAAGT ATTACCTATGAATGAA**TGA**CAAAAGAATCTTCTGGCTAGGGTGTTAGATATATTTATG CATTTTTGGTTTTGTTTTTAAATCAAGCACATCAACCTCAAGCCCGTTTAGCAATGAG
	ORF Start: ATG at 90 **ORF Stop: TGA at 1641**
	SEQ ID NO: 132 **517 aa** **MW at 57179.9kD**
NOV41b, CG57814-02 Protein Sequence	MTCPDKPGQLINWFICSLCVPRVRKLWSSRRPRTRRNLLLGTACAIYLGFLVSQVGRA SLQHGQAAEKGPHRSRDTAEPSFPEIPLDGTLAPPESQGNGSTLQPNVVYITLRSKRS KPANIRGTVKPKRRKKHAVASAAQGQEALVGPSLQPQEAAREADAVALGTLRSKLVKM ESDPEGGAGSGVRAGGPDFLQPSSRESNIRIYSESAPSWLSKDDIRRMRLLADSAVAG LRPVSSRSGARLLVLEGGAPGAVLRCGPSPCGLLKQPLDMSEVFAFHLDRILGLNRTL PSVSRKAEFIQDGRPCPIILWDASLSSASNDTHSSVKLTWGTYQQLLKQKCWQNGRVP KPESGCTEIHHHEWSKMALFDFLLQIYNRLDTNCCGFRPRKEDACVQNGLRPKCDDQG SAALAHIIQRKHDPRHLVFIDNKGFFDRSEDNLNFKLLEGIKEFPASAVSVLKSQHLR QKLLQSLFLDKVYWESQGGRQGIEKLIDVIEHRAKILITYINAHGVKVLPMNE

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 41B.

Table 41B. Comparison of NOV41a against NOV41b.		
Protein Sequence	**NOV41a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV41b	1..519 1..517	493/519 (94%) 497/519 (94%)

Further analysis of the NOV41a protein yielded the following properties shown in Table 41C.

Table 41C. Protein Sequence Properties NOV41a	
PSort analysis:	0.5500 probability located in endoplasmic reticulum (membrane); 0.2404 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Likely cleavage site between residues 59 and 60

A search of the NOV41a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 41D.

223

Table 41D. Geneseq Results for NOV41a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV41a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU12276	Human PRO6001 polypeptide sequence - Homo sapiens, 519 aa. [WO200140466-A2, 07-JUN-2001]	1..519 1..519	518/519 (99%) 519/519 (99%)	0.0
AAM39125	Human polypeptide SEQ ID NO 2270 - Homo sapiens, 519 aa. [WO200153312-A1, 26-JUL-2001]	1..519 1..519	518/519 (99%) 519/519 (99%)	0.0
AAM40911	Human polypeptide SEQ ID NO 5842 - Homo sapiens, 537 aa. [WO200153312-A1, 26-JUL-2001]	1..519 19..537	491/527 (93%) 495/527 (93%)	0.0
AAM41373	Human polypeptide SEQ ID NO 6304 - Homo sapiens, 479 aa. [WO200153312-A1, 26-JUL-2001]	212..512 161..471	130/316 (41%) 180/316 (56%)	1e-64
AAM39587	Human polypeptide SEQ ID NO 2732 - Homo sapiens, 397 aa. [WO200153312-A1, 26-JUL-2001]	212..512 79..389	130/316 (41%) 180/316 (56%)	1e-64

In a BLAST search of public sequence databases, the NOV41a protein was found to have homology to the proteins shown in the BLASTP data in Table 41E.

Table 41E. Public BLASTP Results for NOV41a				
Protein Accession Number	Protein/Organism/Length	NOV41a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9ET25	HYPOTHETICAL BASIC PROTEIN I-19 - Mus musculus (Mouse), 517 aa.	1..519 1..517	431/519 (83%) 462/519 (88%)	0.0
Q9NYZ0	AD021 PROTEIN - Homo sapiens (Human), 246 aa.	274..519 1..246	246/246 (100%) 246/246 (100%)	e-145
Q9UFP1	HYPOTHETICAL 49.5 KDA PROTEIN - Homo sapiens (Human), 448 aa (fragment).	212..512 130..440	129/316 (40%) 179/316 (55%)	2e-63

224

PFam analysis predicts that the NOV41a protein contains the domains shown in the Table 41F.

Table 41F. Domain Analysis of NOV41a			
Pfam Domain	NOV41a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SQS_PSY: domain 1 of 1	109..145	8/37 (22%) 29/37 (78%)	9.9

Example 42.

The NOV42 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 42A.

Table 42A. NOV42 Sequence Analysis		
	SEQ ID NO: 133	1294 bp
NOV42a, CG59327-01 DNA Sequence	GATGGCCACCTTGAACGTTGTACTGATGTTGATGCCCCTTGCCCAGTACATTTTCCAT TGTTTTATAACTGTGCTACTGAAGTACTTGTGTGCAGAGTATGGCTGGAGGAATGCCA TGTTGATCCAAGGCGCCGTTTCCTTAAACCTGTTTGTTTTTGGGACCCTCATGAGGCC CCTCCCTCCTGGGAAAAACCCAAATGACCCAGAAGAGAAAGATCTGCGCGTCCTGCCC GCGCACTCCACAGAGTCTGTAATGTCAAATGGACAGCAGGGAAGAATAGAAGAGAAGG ATGGCGGGTCTGGGAACGAGGAGACCCTCTGTGACCTGCAAGCCCAGGAGTGCAAGCC CAGGAGTGCCCCGATCAGGCCAGATCATGTGCGCTTTCCGGTTCTGAAGACGGTCAGC TGGCTCATTATGAGAGTCAAGAAGGGCTTCGAGGATTGGTACTCAGGCTATTTTGGGA CAGCCAGCCTATTTACAAATCGAATGTTTGTAGCCTTTGTTTTCTGGGCTTCATTTGC ATACAGCAGCTTTGTCATCTCCTTTATTCATCTCCCAGAAATCGTCAATTTGTATAAC TTATTGGAGCAAACGAAGGTTTTCCCTCTGACTTCAATTATAGCAATAGTTCACATTG TTGGAAAAGTGATCCTGGGCGTCATAGCTGACTTACCTTGCATCAGTGTTTGGAATGT CTTCCTGTTGGCCAGCTTCGTTCTTGTCCTCAGTATTTTTGTTTTGCTGCCTTTGATG CATATGTACGCTGGCCTGGTGGTCATCTGCACACTGACAGGGTTTTCCAGCGGTTATT TCTCCCTAATGCCCATAGTGACTGAAGACTTGGTTGGCATTGAACATTTGGCCAATGC CTACGGCATCATCATCTGTGCTAATGGCATCTCTGCGTTGTTGGGACCACCTTTTGCA GGTAAACTGTCTGAGGTTTTAAGAGTTCATAGTGCATATAGATACGGTGTGTTAGCTC TGCGAGGAGACGGATGCAGAGCACTCACATCTTCTCTTATACATAGAAGTGAAATGGC TTTCTAAAGTTAGATCACTGGCCAGAGTTTTTGAGTCACAAGAGCTATTCCACAGATT TCCTTTAGAAAAACAATCACCACTGGCAGTCCACTTCAGTGACACAGAATGGGTTGCA GAACTTGCTTACTTATGTGACACATTCAACCTGCTCAATGAACTCAATCTGTCACTTC AGGGGAGAAGGACAACTGTGTTCAAGTCAGCAAATAAAGTGGCTACATTCAAAACCAA ACTGGAATTACGGGGGTG
	ORF Start: ATG at 2	ORF Stop: TAA at 1049
	SEQ ID NO: 134	349 aa	MW at 38694.2kD
NOV42a, CG59327-01 Protein Sequence	MATLNVVLMLMPLAQYIFHCFITVLLKYLCAEYGWRNAMLIQGAVSLNLFVFGTLMRP LPPGKNPNDPEEKDLRVLPAHSTESVMSNGQQGRIEEKDGGSGNEETLCDLQAQECKP RSAPIRPDHVRFPVLKTVSWLIMRVKKGFEDWYSGYFGTASLFTNRMFVAFVFWASFA YSSFVISFIHLPEIVNLYNLLEQTKVFPLTSIIAIVHIVGKVILGVIADLPCISVWNV FLLASFVLVLSIFVLLPLMHMYAGLVVICTLTGFSSGYFSLMPIVTEDLVGIEHLANA YGIIICANGISALLGPPFAGKLSEVLRVHSAYRYGVLALRGDGCRALTSSLIHRSEMA F

Further analysis of the NOV42a protein yielded the following properties shown in Table 42B.

Table 42B. Protein Sequence Properties NOV42a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 32 and 33

A search of the NOV42a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 42C.

Table 42C. Geneseq Results for NOV42a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAO07132	Human polypeptide SEQ ID NO 21024 - Homo sapiens, 107 aa. [WO200164835-A2, 07-SEP-2001]	257..331 5..81	49/77 (63%) 58/77 (74%)	6e-20
AAY31642	Human transport-associated protein-4 (TRANP-4) - Homo sapiens, 465 aa. [WO9941373-A2, 19-AUG-1999]	157..342 221..401	54/197 (27%) 86/197 (43%)	1e-07
AAY02737	Human secreted protein encoded by gene 88 clone HKAFB88 - Homo sapiens, 229 aa. [WO9902546-A1, 21-JAN-1999]	198..342 24..164	41/147 (27%) 65/147 (43%)	9e-06

In a BLAST search of public sequence databases, the NOV42a protein was found to have homology to the proteins shown in the BLASTP data in Table 42D.

Table 42D. Public BLASTP Results for NOV42a				
Protein Accession Number	Protein/Organism/Length	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96N17	CDNA FLJ30794 FIS, CLONE FEBRA2001093, WEAKLY SIMILAR TO MONOCARBOXYLATE	22..331 1..310	250/312 (80%) 266/312 (85%)	e-138

226

	(Human), 336 aa.			
Q9D1K0	1110004H10RIK PROTEIN - Mus musculus (Mouse), 336 aa.	22..331 1..310	220/312 (70%) 250/312 (79%)	e-119
AAL39716	LD30953P - Drosophila melanogaster (Fruit fly), 894 aa.	142..314 665..843	50/180 (27%) 89/180 (48%)	2e-15
Q9V9B3	CG3409 PROTEIN - Drosophila melanogaster (Fruit fly), 800 aa.	142..314 571..749	50/180 (27%) 89/180 (48%)	2e-15
Q9W0L6	CG13907 PROTEIN - Drosophila melanogaster (Fruit fly), 816 aa.	157..331 565..738	55/178 (30%) 91/178 (50%)	1e-14

PFam analysis predicts that the NOV42a protein contains the domains shown in the Table 42E.

Table 42E. Domain Analysis of NOV42a			
Pfam Domain	NOV42a Match Region	Identities/ Similarities for the Matched Region	Expect Value
oxidored_q3: domain 1 of 1	197..314	25/177 (14%) 73/177 (41%)	9.1

Example 43.

The NOV43 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 43A.

Table 43A. NOV43 Sequence Analysis		
	SEQ ID NO: 135	455 bp
NOV43a, CG59494-01 DNA Sequence	TAGAACTGTGTTGAGCTCTCACCCATCACGATGAGCAACAAATTCTTGGGAACCTGGA AGCTGGTCTCCAGTGAAAACTTTGAGGATTACATGAAAGAACTGGGAGTGAATTTCGC AGCCCGGAACATGGCAGGGTTAGTGAAACCGACAGTAACTATTAGTGTTGATGGGAAA ATGATGACCATAAGAACAGAAAGTTCTTTCCAGGACACTAAGATCTCCTTCAAGCTGG GGGAAGAATTTGATGAAACTACAGCAGACAACCGGAAAGTAAAGAGCACCATAACATT AGAGAATGGCTCAATGATTCACGTCCAAAAATGGCTTGGCAAAGAGACAACAATCAAA AGAAAAATTGTGGATGAAAAAATGGTAGTGGAATGTAAAATGAATAATATTGTCAGCA CCAGAATCTACGAAAAGGTCTGAAAAATCATTTCTTCATTGAAGTGGCT	
	ORF Start: ATG at 31	ORF Stop: TGA at 427
	SEQ ID NO: 136	132 aa	MW at 15096.4kD
NOV43a, CG59494-01 Protein Sequence	MSNKFLGTWKLVSSENFEDYMKELGVNFAARNMAGLVKPTVTISVDGKMMTIRTESSF QDTKISFKLGEEFDETTADNRKVKSTITLENGSMIHVQKWLGKETTIKRKIVDEKMVV ECKMNNIVSTRIYEKV	

Further analysis of the NOV43a protein yielded the following properties shown in Table 43B.

227

Table 43B. Protein Sequence Properties NOV43a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0053 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV43a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 43C.

Table 43C. Geneseq Results for NOV43a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW40227	Human myelin P2 protein - Homo sapiens, 136 aa. [WO9803647-A2, 29-JAN-1998]	1..130 1..130	89/130 (68%) 107/130 (81%)	2e-47
AAW40228	Bovine myelin P2 protein - Bos taurus, 136 aa. [WO9803647-A2, 29-JAN-1998]	1..130 1..130	89/130 (68%) 106/130 (81%)	9e-47
AAY90320	Human AFABP protein sequence - Homo sapiens, 132 aa. [WO200047734-A1, 17-AUG-2000]	1..131 1..131	84/131 (64%) 110/131 (83%)	3e-46
AAY90319	Mouse AFABP protein sequence - Mus sp, 132 aa. [WO200047734-A1, 17-AUG-2000]	1..131 1..131	83/131 (63%) 108/131 (82%)	7e-45
AAG66576	Mouse MDGI polypeptide - Mus sp, 133 aa. [US6232291-B1, 15-MAY-2001]	1..131 1..131	73/131 (55%) 103/131 (77%)	6e-40

In a BLAST search of public sequence databases, the NOV43a protein was found to have homology to the proteins shown in the BLASTP data in Table 43D.

228

Table 43D. Public BLASTP Results for NOV43a				
Protein Accession Number	Protein/Organism/Length	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
MPRB2	myelin P2 protein - rabbit, 132 aa.	1..132 1..132	95/132 (71%) 109/132 (81%)	3e-49
P02691	Myelin P2 protein - Oryctolagus cuniculus (Rabbit), 131 aa.	2..132 1..131	94/131 (71%) 108/131 (81%)	1e-48
MPHU2	myelin P2 protein [validated] - human, 132 aa.	1..132 1..132	92/132 (69%) 109/132 (81%)	3e-48
Q90X56	ADIPOCYTE FATTY ACID BINDING PROTEIN - Gallus gallus (Chicken), 132 aa.	1..131 1..131	86/131 (65%) 113/131 (85%)	1e-47
P02689	Myelin P2 protein - Homo sapiens (Human), 131 aa.	2..132 1..131	91/131 (69%) 108/131 (81%)	1e-47

PFam analysis predicts that the NOV43a protein contains the domains shown in the Table 43E.

Table 43E. Domain Analysis of NOV43a			
Pfam Domain	NOV43a Match Region	Identities/ Similarities for the Matched Region	Expect Value
lipocalin: domain 1 of 1	4..132	45/157 (29%) 113/157 (72%)	3.2e-36

Example 44.

The NOV44 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 44A.

Table 44A. NOV44 Sequence Analysis		
	SEQ ID NO: 137	1561 bp
NOV44a, CG59432-01 DNA Sequence	AAGAATTGTAGCTCTCCACTGAATTGCAGGGGTTCTTGAATGTTGTCAACATTTGGAG GCAGTTGGAGGAGGGAGCTCTATTGATGAAAAATGGCTACATATTCAAAATTTCAGTG TATACCAGGAAGATAATTCAATTCAATCTCTGGCTTACCCAAAGAATCTTGGAGTTAC TGCCAATGAGGAAATCCCCAGGGTCTAATAAAAATATCTTTAGGAGTGAAGGAGTTAA CTGAGTGTGTAAGCTTTATCTTCTGTCCAATGGACTTGTGGTTTGCTTATAAAACTCT CCAGTAAATAATTGTTAGAGACCTGTCATTGATAGCAGTTGCTAGTTGCTGCCTTTTA AGAGCTCGTTGATTCCTCTGCAAGGTGGTGCAGCATCCTCTGTCCCTTCATTCATTTC AGATCTACTCAGGTCTCCCTGTAAACAGATCTCTCGGATCAATAAGCATGAATGACGA AGACTACAGCACCATCTATGACACAATCCAAAATGAGAGGACGTATGAGGTTCCAGAC CAGCCAGAAGAAAATGAAAGTCCCCATTATGATGATGTCCATGAGTACTTAAGGCCAG AAAATGATTTATATGCCACTCAGCTGAATACCCATGAGTATGATTTTGTGTCAGTCTA	

229

	TACCATTAAGGGTGAAGAGACCAGCTTGGCCTCTGTCCAGTCAGAAGACAGAGGCTAC CTCCTGCCTGATGAGATATACTCTGAACTCCAGGAGGCTCATCCAGGTGAGCCCCAGG AGGACAGGGGCATCTCAATGGAAGGGTTATATTCATCAACCCAGGACCAGCAACTCTG CGCAGCAGAACTCCAGGAGAATGGGAGTGTGATGAAGGAAGATCTGCCTTCTCCTTCA AGCTTCACCATTCAGCACAGTAAGGCCTTCTCTACCACCAAGTATTCCTGCTATTCTG ATGCTGAAGGTTTGGAAGAAAAGGAGGGAGCTCACATGAACCCTGAGATTTACCTCTT TGTGAAGGTAAGGTCTGCCTCTGACAGGCATACCCTGTTCATGCAGATATTATGGCTG GTGTTTTATTTTGCTCTGAATGACCAGGGAAAGATTCATAATGCCATGGTCCTTGGAT CTCAATACATATTCAGGAGTCGGAGGGAC**TAA**ATCAGTCATTAGAGTGTACTCAGCTC TTCACAAAATTAGAGGAATTGGAAGGTGCATTTAAAGCACGTATTTAATCACTGACTT TTACATACCATGGGCAAAGTATTTTTCAAAACGGTTCACATAAGTGAGCCATAACTGC TGCCCAAATCCTTGCCATTGTGGCTGACATTAAGTACATTTTTCTGTCTGGTTAAATT TCCTTTGTCGACATGTTTAAAAGTGAAACCAAAGCTTGTGAAAGAAAGACCTTCTTGT GCTTCTAAGGTCACAGATTTGTCAGATAGGTGGTCAATAAAGGCTATCTCTGTCACTA GCTTGCCCCTTTGGCACCAATATAACTAAAAATTTGATGAAGTCAAATGATTTCAGTA GTAGTAAGACACTACCAGTGTTAATGTTTAATACTTACGATATCTAAACAGAA
	ORF Start: ATG at 454 **ORF Stop: TAA at 1132**
	SEQ ID NO: 138 **226 aa** **MW at 26132.2kD**
NOV44a, **CG59432-01 Protein Sequence**	MNDEDYSTIYDTIQNERTYEVPDQPEENESPHYDDVHEYLRPENDLYATQLNTHEYDF VSVYTIKGEETSLASVQSEDRGYLLPDEIYSELQEAHPGEPQEDRGISMEGLYSSTQD QQLCAAELQENGSVMKEDLPSPSSFTIQHSKAFSTTKYSCYSDAEGLEEKEGAHMNPE IYLFVKVRSASDRHTLFMQILWLVFYFALNDQGKIHNAMVLGSQYIFRSRRD
	SEQ ID NO: 139 **809 bp**
NOV44b, **CG59432-02 DNA Sequence**	ATCCTCTGTCCCTTCATTCATTTCAGATCTACTCAGGTCTCCCTGTAAACAGATCTCT CGGATCAATAAGC**ATG**AATGACGAAGACTACGGCACCATCTATGACACAATCCAAAAT GAGAGGACGTATGAGGTTCCAGACCAGCCAGAAGAAAATGAAAGTCCCCATTATGATG ATGTCCATGAGTACTTAAGGCCAGAAAATGATTTATATGCCACTCAGCTGAATACCCA TGAGTATGATTTTGTGTCAGTCTATACCATTAAGGGTGAAGAGACCAGCTTGGCCTCT GTCCAGTCAGAAGACAGAGGCTACCTCCTGCCTGATGAGATATACTCTGAACTCCAGG AGGCTCATCCAGGTGAGCCCCAGGAGGACAGGGGCATCTCAATGGAAGGGTTATATTC ATCAACCCAGGACCAGCAACTCTGCGCAGCAGAACTCCAGGAGAATGGGAGTGTGATG AAGGAAGATCTGCCTTCTCCTTCAAGCTTCACCATTCAGCACAGTAAGGCCTTCTCTA CCACCAAGTATTCCTGCTATTCTGATGCTGAAGGTTTGGAAGAAAAGGAGGGAGCTCA CATGAACCCTGAGATTTACCTCTTTGTGAAGGTAAGGTCTGCCTCTGACAGGCATACC CTGTTCATGCAGATATTATGGCTGGTGTTTTATTTTGCTCTGAATGACCAGGGAAAGA TTCATAATGCCATGGTCCTTGGATCTCAATACATATTCAGGAGTCGGAGGGAC**TAA**AT CAGTCATTAGAGTGTACTCAGCTCTTCACAAAATTAGAGGAATTGGAAGGTGCAT
	ORF Start: ATG at 72 **ORF Stop: TAA at 750**
	SEQ ID NO: 140 **226 aa** **MW at 26102.2kD**
NOV44b, **CG59432-02 Protein Sequence**	MNDEDYGTIYDTIQNERTYEVPDQPEENESPHYDDVHEYLRPENDLYATQLNTHEYDF VSVYTIKGEETSLASVQSEDRGYLLPDEIYSELQEAHPGEPQEDRGISMEGLYSSTQD QQLCAAELQENGSVMKEDLPSPSSFTIQHSKAFSTTKYSCYSDAEGLEEKEGAHMNPE IYLFVKVRSASDRHTLFMQILWLVFYFALNDQGKIHNAMVLGSQYIFRSRRD

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 44B.

Table 44B. Comparison of NOV44a against NOV44b.		
Protein Sequence	**NOV44a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV44b	1..226 1..226	225/226 (99%) 225/226 (99%)

Further analysis of the NOV44a protein yielded the following properties shown in Table 44C.

Table 44C. Protein Sequence Properties NOV44a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV44a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 44D.

Table 44D. Geneseq Results for NOV44a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV44a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found				

In a BLAST search of public sequence databases, the NOV44a protein was found to
5   have homology to the proteins shown in the BLASTP data in Table 44E.

Table 44E. Public BLASTP Results for NOV44a				
Protein Accession Number	Protein/Organism/Length	NOV44a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96JT5	CLIC5B - Homo sapiens (Human), 410 aa.	1..200 1..202	185/202 (91%) 191/202 (93%)	e-104
Q9NPY9	DJ447E21.4 (SIMILAR TO BOVINE CHLORIDE CHANNEL PROTEIN (P64)) - Homo sapiens (Human), 180 aa (fragment).	1..180 1..180	180/180 (100%) 180/180 (100%)	e-103
A47104	chloride channel 64K chain - bovine, 437 aa.	1..197 1..229	104/231 (45%) 133/231 (57%)	1e-39
P35526	Chlorine channel protein p64 - Bos taurus (Bovine), 437 aa.	1..197 1..229	103/231 (44%) 131/231 (56%)	1e-38

PFam analysis predicts that the NOV44a protein contains the domains shown in the Table 44F.

231

Table 44F. Domain Analysis of NOV44a			
Pfam Domain	NOV44a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 45.

The NOV45 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 45A.

Table 45A. NOV45 Sequence Analysis		
	SEQ ID NO: 141	877 bp
NOV45a, CG59394-01 DNA Sequence	ACTTTGTCCTCTTGGGCTTCACACAGAATCCAAAGGAGCAGAAAGTACTTTTTGTTAT GTTCTTGCTCTTCTACATTTTGACCATGGTGGGCAACCTGCTCATTGTAGTGACCGTA ACTGTCAGTGAGACCCTGGGCTCACCAATGTACTTCTTTCTTGCTGGCTTATCATTTA TAGATATCATTTATTCTTCATCCATTTCCCCCAGATTGATTTCAGGCTTGTTCTTTGG GAATAATTCCATATCCTTCCAATCTTGCATGGCCCAGCTCTTTATCGAGCACATTTTC GGTGGGTCAGAGGTCTTTCTCCTGTTGGTGATGGCCTATGACTGCTATGTGGCCATCT GTAAGCCCTTGCATTATTTGGTTATCATGAGACAATGGGTGTGTGTTGTGCTGCTGGT AGTGTCCTGGGTTGGAGGATTTCTGCACTCAGTATTTCAACTTAGCATTATTTATGGG CTCCCATTCTGTGGCCCCAATGTCATTGATCATTTTTTCTGTGACATGTATCCCTTAT TGAAACTGGTCTGCACTGACACCCATGCTATTGGCCTCTTAGTGGTGGCCAATGGAGG ACTGGCTTGCACTATTGTGTTTCTGCTCTTACTCATCTCTTATGGTGTCATCTTGCAC TCTTTAAAGAACCTTAGTCAGAAAGGGAGGCAAAAAGCCCTCTCAACCTGCAGTTCCC ACATGACTGTGGTTGTCTTCTTCTTTGTTCCTTGTATTTTTATGTATGCTAGACCTGC TAGGACCTTCCCCATTGACAAATCAGTGAGTGTGTTTTATACAGTCATAACCCCAATG CTGAACCCCTTAATCTACACTCTGAGAAATTCTGAGATGACAAGTGCTATGAAGAAGC TTTAGAG
	ORF Start: TTT at 3	ORF Stop: TAG at 873
	SEQ ID NO: 142	290 aa	MW at 32485.7kD
NOV45a, CG59394-01 Protein Sequence	FVLLGFTQNPKEQKVLFVMFLLFYILTMVGNLLIVVTVTVSETLGSPMYFFLAGLSFI DIIYSSSISPRLISGLFFGNNSISFQSCMAQLFIEHIFGGSEVFLLLVMAYDCYVAIC KPLHYLVIMRQWVCVVLLVVSWVGGFLHSVFQLSIIYGLPFCGPNVIDHFFCDMYPLL KLVCTDTHAIGLLVVANGGLACTIVFLLLLISYGVILHSLKNLSQKGRQKALSTCSSH MTVVVFFFVPCIFMYARPARTFPIDKSVSVFYTVITPMLNPLIYTLRNSEMTSAMKKL

Further analysis of the NOV45a protein yielded the following properties shown in Table 45B.

Table 45B. Protein Sequence Properties NOV45a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 42 and 43

A search of the NOV45a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 45C.

Table 45C. Geneseq Results for NOV45a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV45a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU24536	Human olfactory receptor AOLFR21 - Homo sapiens, 299 aa. [WO200168805-A2, 20-SEP-2001]	1..290 10..299	273/290 (94%) 278/290 (95%)	e-155
AAG71950	Human olfactory receptor polypeptide, SEQ ID NO: 1631 - Homo sapiens, 299 aa. [WO200127158-A2, 19-APR-2001]	1..290 10..299	273/290 (94%) 278/290 (95%)	e-155
AAG72258	Human olfactory receptor polypeptide, SEQ ID NO: 1939 - Homo sapiens, 262 aa. [WO200127158-A2, 19-APR-2001]	33..290 1..250	234/258 (90%) 240/258 (92%)	e-131
AAG72553	Human OR-like polypeptide query sequence, SEQ ID NO: 2234 - Homo sapiens, 327 aa. [WO200127158-A2, 19-APR-2001]	1..290 10..299	198/290 (68%) 242/290 (83%)	e-121
AAG71909	Human olfactory receptor polypeptide, SEQ ID NO: 1590 - Homo sapiens, 327 aa. [WO200127158-A2, 19-APR-2001]	1..290 10..299	198/290 (68%) 242/290 (83%)	e-121

In a BLAST search of public sequence databases, the NOV45a protein was found to have homology to the proteins shown in the BLASTP data in Table 45D.

Table 45D. Public BLASTP Results for NOV45a				
Protein Accession Number	Protein/Organism/Length	NOV45a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9QW37	OR18=ODORANT RECEPTOR - Rattus sp, 307 aa.	1..290 10..299	192/290 (66%) 237/290 (81%)	e-115
Q96R66	OLFACTORY RECEPTOR - Homo sapiens (Human), 213 aa (fragment).	57..269 1..213	198/213 (92%) 202/213 (93%)	e-111
Q9R0K2	ODORANT RECEPTOR MOR18 - Mus musculus (Mouse), 308 aa.	1..290 10..299	177/290 (61%) 229/290 (78%)	e-105

Q9R0K1	ODORANT RECEPTOR A16 - Mus musculus (Mouse), 302 aa.	1..290 10..299	171/290 (58%) 226/290 (76%)	e-102
CAC88333	SEQUENCE 34 FROM PATENT WO0164879 - Homo sapiens (Human), 309 aa.	1..290 10..299	167/290 (57%) 221/290 (75%)	5e-99

PFam analysis predicts that the NOV45a protein contains the domains shown in the Table 45E.

<table>
<tr><th colspan="4">Table 45E. Domain Analysis of NOV45a</th></tr>
<tr><th>Pfam Domain</th><th>NOV45a Match Region</th><th>Identities/ Similarities for the Matched Region</th><th>Expect Value</th></tr>
<tr><td>7tm_1: domain 1 of 1</td><td>30..276</td><td>50/268 (19%) 174/268 (65%)</td><td>4.4e-23</td></tr>
</table>

Example 46.

The NOV46 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 46A.

<table>
<tr><th colspan="3">Table 46A. NOV46 Sequence Analysis</th></tr>
<tr><td></td><td>SEQ ID NO: 143</td><td>1746 bp</td></tr>
<tr><td>NOV46a,
CG59383-01 DNA Sequence</td><td colspan="2">ATAATTCAGTTTGAAAACCAGTGGTTTCTCTTTCCTTCCCTATAGGTGTAAAGAATAT
CCAGCTGGTGGCTACAGTTCCCCCTCTGGTTTTGCTGCCATGCATCCTGGGCGAACTA
CTGGTAAAGGGCCCTCTACTCACACTCAGATTGACCAGCAACCTCCACGGCTTCTCAT
TGTGCACATTGCTCTACCGTCCTGGGCTGACATCTGCACCAACCTCTGTGAGGCTCTG
CAGAACTTCTTCTCTCTAGCCTGCAGCTTGATGGGCCCCAGCCGCATGTCCCTGTTCA
GTTTATACATGGTACAAGATCAGCATGAGTGCATCCTCCCTTTTGTGCAAGTGAAAGG
GAACTTTGCTAGGTTGCAGACCTGCATCTCAGAACTCCGCATGTTACAGAGAGAAGGG
TGTTTCAGATCACAAGGTGCTTCTCTGCGGCTGGCAGTAGAGGATGGGCTCCAGCAAT
TCAAACAATACAGCAGACATGTGACCACAAGGGCAGCTCTGACCTATACCTCCCTGGA
GATTACTATTCTGACTTCTCAGCCTGGAAAAGAGGTGGTCAAACAGTTGGAGGAAGGG
TTGAAAGATACAGACCTAGCCAGAGTCAGGAGGTTTCAGGTCGTTGAGGTCACAAAGG
GAATCCTAGAGCACGTGGACTCAGCGTCTCCTGTTGAGGATACCAGCAATGATGAGAG
TTCTATTCTGGGAACTGACATTGACCTTCAGACTATAGACAATGATATCGTCAGCATG
GAGATTTTCTTCAAAGCCTGGCTACATAACAGTGGAACAGACCAAGAACAAATCCATC
TTCTTCTTTCTTCACAGTGTTTCAGCAACATTTCCAGACCCAGAGATAATCCAATGTG
TCTGAAATGTGATCTCCAAGAGCGACTGCTCTGCCCATCCCTACTCGCTGGCACAGCT
GACGGCTCCTTGAGAATGGATGACCCTAAAGGAGACTTCATCACACTCTACCAGATGG
CTTCCCAGTCATCGGCCTCTCATTACAAGCTCCAAGTGATCAAGGCTTTAAAATCTAG
CGGGCTCTGCGAGTCATTGACATATGGACTCCCGTTCATCCTCAGACCTACAAGCTGT
TGGCAGCTGGACTGGGATGAGCTGGAGACAAATCAGCAACATTTCCATGCTTTGTGTC
ACAGCCTGCTGAAAAGGGAATGGCTGCTGTTAGCCAAGGGGGAACCACCGGGCCCAGG
ACACAGCCAGAGAATTCCTGCCAGCACCTTCTATGTGATCATGCCGTCACACTCCCTC
ACACTGCTGGTAAAGGCGGTGGCCACGCGGGAACTGATGCTGCCCAGCACCTTCCCCC
TGCTACCTGAGGACCCCACATGATGATAGCCTTAAGAATAGCATGCTGGACAGCCTGGA
GCTGGAGCCCACCTACAACCCCTTGCATGTTCAAAGCCACCTGTACTCACACCTGAGC
AGCATCTATGCCAAGCCTCAGGGGCGGCTCCACCCACACTGGGAGAGCCGAGCTCCGA
GAAAGACTGGGCAGTTGCAGACCAACCGAGCTCGAGCTACTGTGGCCCCCCTGCCTAT
GACTCCTGTCCCAGGCAGAGCCTCCAAGATGCCAGCAGCCAGCAAATCTTCCTCAGAT
GCCTTCTTCCTGCCTTCAGAGTGGGAGAAGGATCCCTCAAGGCCCTAAGTCACCAGCA
CCAGAGCCCAGCTGCCCAGCTTAACCATATCCATGCTCAGGTTCACATAATGGCTATC
TGTGGT</td></tr>
<tr><td></td><td>ORF Start: ATG at 98</td><td>ORF Stop: TAA at 1670</td></tr>
<tr><td></td><td>SEQ ID NO: 144</td><td>524 aa | MW at 58691.3kD</td></tr>
</table>

234

NOV46a, CG59383-01 Protein Sequence	MHPGRTTGKGPSTHTQIDQQPPRLLIVHIALPSWADICTNLCEALQNFFSLACSLMGP SRMSLFSLYMVQDQHECILPFVQVKGNFARLQTCISELRMLQREGCFRSQGASLRLAV EDGLQQFKQYSRHVTTRAALTYTSLEITILTSQPGKEVVKQLEEGLKDTDLARVRRFQ VVEVTKGILEHVDSASPVEDTSNDESSILGTDIDLQTIDNDIVSMEIFFKAWLHNSGT DQEQIHLLLSSQCFSNISRPRDNPMCLKCDLQERLLCPSLLAGTADGSLRMDDPKGDF ITLYQMASQSSASHYKLQVIKALKSSGLCESLTYGLPFILRPTSCWQLDWDELETNQQ HFHALCHSLLKREWLLLAKGEPPGPGHSQRIPASTFYVIMPSHSLTLLVKAVATRELM LPSTFPLLPEDPHDDSLKNSMLDSLELEPTYNPLHVQSHLYSHLSSIYAKPQGRLHPH WESRAPRKTGQLQTNRARATVAPLPMTPVPGRASKMPAASKSSSDAFFLPSEWEKDPS RP
	SEQ ID NO: 145 1647 bp
NOV46b, CG59383-02 DNA Sequence	AAAGAATATCCAGCTGGTGGCTACAGTTCCCCCTCTGGTTTTGCTGCC**ATG**CATCCTG GGCGAACTACTGGTAAAGGGCCCTCTACTCACACTCAGATTGACCAGCAACCTCCACG GCTTCTCATTGTGCACATTGCTCTACCGTCCTGGGCTGACATCTGCACCAACCTCTGT GAGGCTCTGCAGAACTTCTTCTCTCTAGCCTGCAGCTTGATGGGCCCCAGCCGCATGT CCCTGTTCAGTTTATACATGGTACAAGATCAGCATGAGTGCATCCTCCCTTTTGTGCA AGTGAAAGGGAACTTTGCTAGGTTGCAGACCTGCATCTCAGAACTCCGCATGTTACAG AGAGAAGGGTGTTTCAGATCACAAGGTGCTTCTCTGCGGCTGGCAGTAGAGGATGGGC TCCAGCAATTCAAACAATACAGCAGACATGTGACCACAAGGGCAGCTCTGACCTATAC CTCCCTGGAGATTACTATTCTGACTTCTCAGCCTGGAAAAGAGGTGGTCAAACAGTTG GAGGAAGGGTTGAAAGATACAGACCTAGCCAGAGTCAGGAGGTTTCAGGTCGTTGAGG TCACAAAGGGAATCCTAGAGCACGTGGACTCAGCGTCTCCTGTTGAGGATACCAGCAA TGATGAGAGTTCTATTCTGGGAACTGACATTGACCTTCAGACTATAGACAATGATATC GTCAGCATGGAGATTTTCTTCAAAGCCTGGCTACATAACAGTGGAACAGACCAGGAAC AAATCCATCTTCTTCTTTCTTCACAGTGTTTCAGCAACATTTCCAGACCCAGAGATAA TCCAATGTGTCTGAAATGTGATCTCCAAGAGCGACTGCTCTGCCCATCCCTACTCGCT GGCACAGCTGACGGCTCCTTGAGAATGGATGACCCTAAAGGAGACTTCATCACACTCC ACCAGATGGCTTCCCAGTCATCGGCCTCTCATTACAAGCTCCAAGTGATCAAGGCTTT AAAATCTAGCGGGCTCTGCGAGTCATTGACATATGGACTCCCGTTCATCCTCAGACCT ACAAGCTGTTGGCAGCTGGACTGGGATGAGCTGGAGACAAATCAGCAACATTTCCATG CTTTGTGTCACAGCCTGCTGAAAAGGGAATGGCTGCTGTTAGCCAAGGGGGAACCACC GGGCCCAGGACACAGCCAGAGAATTCCTGCCAGCACCTTCTATGTGATCATGCCGTCA CACTCCCTCACACTGCTGGTAAAGGCGGTGGCCACGCGGGAACTGATGCTGCCCAGCA CCTTCCCCCTGCTGCCTGAGGACCCACATGATGATAGCCTTAAGAATGTGGAGAGCAT GCTGGACAGCCTGGAGCTGGAGCCCACCTACAACCCCTTGCATGTTCAAAGCCACCTG TACTCACACCTGAGCAGCATCTATGCCAAGCCTCAGGGGCGGCTCCACCCACACTGGG AGAGCCGAGCTCCGAGAAAGCATCCCTGCAAGACTGGGCAGTTGCAGACCAACCGAGC TCGAGCTACTGTGGCCCCCCTGCCTATGACTCCTGTCCCAGGCAGAGCCTCCAAGATG CCAGCAGCCAGCAAATCTTCCTCAGATGCCTTCTTCCTGCCTTCAGAGTGGGAGAAGG ATCCCTCAAGGCCC**TAA**GTCACC
	ORF Start: ATG at 49 ORF Stop: TAA at 1639
	SEQ ID NO: 146 530 aa MW at 59359.1kD
NOV46b, CG59383-02 Protein Sequence	MHPGRTTGKGPSTHTQIDQQPPRLLIVHIALPSWADICTNLCEALQNFFSLACSLMGP SRMSLFSLYMVQDQHECILPFVQVKGNFARLQTCISELRMLQREGCFRSQGASLRLAV EDGLQQFKQYSRHVTTRAALTYTSLEITILTSQPGKEVVKQLEEGLKDTDLARVRRFQ VVEVTKGILEHVDSASPVEDTSNDESSILGTDIDLQTIDNDIVSMEIFFKAWLHNSGT DQEQIHLLLSSQCFSNISRPRDNPMCLKCDLQERLLCPSLLAGTADGSLRMDDPKGDF ITLHQMASQSSASHYKLQVIKALKSSGLCESLTYGLPFILRPTSCWQLDWDELETNQQ HFHALCHSLLKREWLLLAKGEPPGPGHSQRIPASTFYVIMPSHSLTLLVKAVATRELM LPSTFPLLPEDPHDDSLKNVESMLDSLELEPTYNPLHVQSHLYSHLSSIYAKPQGRLH PHWESRAPRKHPCKTGQLQTNRARATVAPLPMTPVPGRASKMPAASKSSSDAFFLPSE WEKDPSRP

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 46B.

Table 46B. Comparison of NOV46a against NOV46b.		
Protein Sequence	NOV46a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV46b	1..524 1..530	509/530 (96%) 510/530 (96%)

Further analysis of the NOV46a protein yielded the following properties shown in Table 46C.

Table 46C. Protein Sequence Properties NOV46a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV46a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 46D.

Table 46D. Geneseq Results for NOV46a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV46a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM34317	Peptide #8354 encoded by probe for measuring placental gene expression - Homo sapiens, 52 aa. [WO200157272-A2, 09-AUG-2001]	259..310 1..52	52/52 (100%) 52/52 (100%)	7e-23
ABB18624	Protein #623 encoded by probe for measuring heart cell gene expression - Homo sapiens, 42 aa. [WO200157274-A2, 09-AUG-2001]	101..142 1..42	42/42 (100%) 42/42 (100%)	2e-16
AAM66343	Human bone marrow expressed probe encoded protein SEQ ID NO: 26649 - Homo sapiens, 42 aa. [WO200157276-A2, 09-AUG-2001]	101..142 1..42	42/42 (100%) 42/42 (100%)	2e-16
AAM53955	Human brain expressed single exon probe encoded protein SEQ ID NO: 26060 - Homo sapiens, 42 aa. [WO200157275-A2, 09-AUG-2001]	101..142 1..42	42/42 (100%) 42/42 (100%)	2e-16
AAM26622	Peptide #659 encoded by probe for measuring placental gene expression - Homo sapiens, 42 aa. [WO200157272-A2, 09-AUG-2001]	101..142 1..42	42/42 (100%) 42/42 (100%)	2e-16

In a BLAST search of public sequence databases, the NOV46a protein was found to have homology to the proteins shown in the BLASTP data in Table 46E.

Table 46E. Public BLASTP Results for NOV46a				
Protein Accession Number	Protein/Organism/Length	NOV46a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Z0E1	D6MM5E PROTEIN - Mus musculus (Mouse), 529 aa.	1..524 1..526	380/526 (72%) 423/526 (80%)	0.0
Q96L07	SIMILAR TO DNA SEGMENT, CHR 6, MIRIAM MEISLER 5, EXPRESSED - Homo sapiens (Human), 365 aa.	1..358 1..358	358/358 (100%) 358/358 (100%)	0.0

PFam analysis predicts that the NOV46a protein contains the domains shown in the
5   Table 46F.

Table 46F. Domain Analysis of NOV46a			
Pfam Domain	NOV46a Match Region	Identities/ Similarities for the Matched Region	Expect Value
RA: domain 1 of 1	124..214	18/115 (16%) 65/115 (57%)	8.4

Example 47.

The NOV47 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 47A.

Table 47A. NOV47 Sequence Analysis		
	SEQ ID NO: 147	960 bp
NOV47a, CG58526-01 DNA Sequence	AGGACTAAATAAAATGGCCTAAATTTAAATATGGATTGGGATTTCCATTCTCTTGCAG ATGCCCAGAACCAAAGAAGAGGTCTGCCTGGTTTTCTTCCTGGAGCTCCAGACCCAGA CCAAAGCCTTCCTGCCTCTTCCAATCCAGGGAACCAAGCATGGCAGCTGAGTCTCCCT CTGCCAAGCAGTTTCCTGCCAACAGTCAGTCTCCCTCCTGGTCTAGAATATTTAAGCC AGTTAGACCTGATAATTATACACCAGCAGGTGGAGCTGCTTGTGATACTTGGTACTGA GACCTCCAACAAATATGAGATTAAAAACAGCTTGGGACAAAGAATTTACTTTGCAGTG GAGGAAAGCATCTGCTTCAATCGTACTTTCTGTTCCACTCTGCGATCTTGCACCCTGA GGATCACAGATAACTCAGGTCGAGAGGTCATTACAGTGAACAGGCCCTTGAGATGTAA CAGCTGCTGGTGCCCTTGCTACCTACAAGAGTTAGAAATCCAAGCCCCTCCTGGTACT ATAGTTGGTTACGTTACGCAGAAGTGGGACCCCTTTCTGCCTAAATTCACAATCCAAA ATGCAAACAAAGAAGATATTTTGAAAATTGTTGGTCCTTGTGTGACATGTGGCTGTTT TGGCGATGTGGATTTTGAGAAGGTGAAAACCATTAATGAAAAGCTTACAATTGGGAAG ATTTCAAAGTACTGGTCAGGATTTGTAAATGATGTCTTCACAAATGCTGACAATTTCG GAATTCATGTTCCTGCAGATCTAGATGTAACAGTCAAAGCAGCAATGATCGGTGCCTG TTTTCTCTTTGTAAGTATGGGCTTTGAGAGCCCAGCCCTCCAAGATGAGAAAGAGTCA GTGTGGCAATTCAAAAAATCAGAGTGCCCTCTCACCTCCAAACAAGCCCACTTGTTCC

237

	CCAGCGATGGTTCTTAGCCAGACTGAAATGAC
	ORF Start: ATG at 31 | ORF Stop: TAG at 943
	SEQ ID NO: 148 | 304 aa | MW at 33794.2kD
NOV47a, CG58526-01 Protein Sequence	MDWDFHSLADAQNQRRGLPGFLPGAPDPDQSLPASSNPGNQAWQLSLPLPSSFLPTVS LPPGLEYLSQLDLIIHQQVELLVILGTETSNKYEIKNSLGQRIYFAVEESICFNRTF CSTLRSCTLRITDNSGREVITVNRPLRCNSCWCPCYLQELEIQAPPGTIVGYVTQKWD PFLPKFTIQNANKEDILKIVGPCVTCGCFGDVDFEKVKTINEKLTIGKISKYWSGFVN DVFTNADNFGIHVPADLDVTVKAAMIGACFLFVSMGFESPALQDEKESVWQFKKSECP LTSKQAHLFPSDGS

Further analysis of the NOV47a protein yielded the following properties shown in Table 47B.

Table 47B. Protein Sequence Properties NOV47a	
PSort analysis:	0.8500 probability located in endoplasmic reticulum (membrane); 0.4400 probability located in plasma membrane; 0.4244 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV47a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 47C.

Table 47C. Geneseq Results for NOV47a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV47a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG78341	Human Mm-1 cell line derived transplantability-associated gene 1b - Homo sapiens, 318 aa. [WO200164894-A2, 07-SEP-2001]	24..282 60..318	152/263 (57%) 187/263 (70%)	5e-84
AAB24113	Human phospholipid scramblase HPLS protein sequence - Homo sapiens, 318 aa. [CN1259574-A, 12-JUL-2000]	24..282 60..318	152/263 (57%) 187/263 (70%)	5e-84
AAB24112	Mouse phospholipid scramblase MPLS protein sequence - Mus sp, 318 aa. [CN1259574-A, 12-JUL-2000]	24..282 60..318	152/263 (57%) 187/263 (70%)	5e-84

238

AAY09309	Human phospholipid scramblase - Homo sapiens, 318 aa. [WO9919352-A2, 22-APR-1999]	24..282 60..318	152/263 (57%) 187/263 (70%)	5e-84
AAY29323	Human PL scramblase - Homo sapiens, 318 aa. [WO9936536-A2, 22-JUL-1999]	24..282 60..318	152/263 (57%) 187/263 (70%)	5e-84

In a BLAST search of public sequence databases, the NOV47a protein was found to have homology to the proteins shown in the BLASTP data in Table 47D.

Table 47D. Public BLASTP Results for NOV47a				
Protein Accession Number	Protein/Organism/Length	NOV47a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9JJ00	Phospholipid scramblase 1 (PL scramblase 1) (Transplantability associated protein 1) (TRA1) (NOR1) - Mus musculus (Mouse), 328 aa.	20..283 66..328	150/267 (56%) 191/267 (71%)	4e-84
Q99M50	PHOSPHOLIPID SCRAMBLASE 1 - Mus musculus (Mouse), 327 aa.	20..282 66..327	150/266 (56%) 191/266 (71%)	6e-84
O15162	Phospholipid scramblase 1 (PL scramblase 1) (Erythrocyte phospholipid scramblase) (Ca2+ dependent phospholipid scramblase 1) (MmTRA1b) - Homo sapiens (Human), 318 aa.	24..282 60..318	152/263 (57%) 187/263 (70%)	2e-83
P58195	Phospholipid scramblase 1 (PL scramblase 1) (Ca2+ dependent phospholipid scramblase 1) - Rattus norvegicus (Rat), 335 aa.	28..282 84..335	145/256 (56%) 183/256 (70%)	3e-81
Q9NRY7	Phospholipid scramblase 2 (PL scramblase 2) (Ca2+ dependent phospholipid scramblase 2) - Homo sapiens (Human), 224 aa.	55..270 6..221	135/217 (62%) 164/217 (75%)	1e-75

PFam analysis predicts that the NOV47a protein contains the domains shown in the

5   Table 47E.

239

Table 47E. Domain Analysis of NOV47a			
Pfam Domain	NOV47a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 48.

The NOV48 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 48A.

Table 48A. NOV48 Sequence Analysis		
	SEQ ID NO: 149	957 bp
NOV48a, CG57851-01 DNA Sequence	CCCCTGCTGGTGCCCAAGACCACCGTGGAAGGAATGGCTAAAGAGGAGACAAGTGAGT TAGAATGGGGCTTGTTACCCCCAGAAGAATTTTCCCAAGTGAATGGAATCATTCTTCA AAAGAAAATGTGCGATTTCTGGGATAAGATCTGGAACTTCCAAGCCAAGCCTGATGAC CTGCTCATTGCTTCTTACCCCAAAGCAGGTACCACTTGGACACAGGAAATTGTAGATC TGATACAAATGATGGCGATATTGAGAAAGCAGGCGCGCTTCCATTCAACTTCAACA CCCTTTCCTGGAGTGGATAAGAATGACACACGCCAGGAAAATTTTTGCAGGGATTGAC CAGGCTAACACAATGCCTTCCCCAAGGACCCTGAAAACTCATCTTCCTGTACAACTAC TGCCTCCATCCTTCTGGGAGGAAAACTGTAAGATAATCTATGTGGCAAGAAATGCCAA GGATAACCTGGTGTCCTACTACCATTTTCAAAGGATGAGCAAAGCACTCCCTGACGTT TTGACAGTGGGAGAATACATTATGTGTGGGGAAGTGTTGTGGGGAATATGGGAAGAGA TTCGGACTTGGCAACTGCATAGGTTGTTCTGCTGGTTCTTTGATCATGCTTCTGAGAA TCCTAGAAAGTTCAAAAGGATAATGGAATTTATGGGGAATAAACTAGATGAAGATCCT GTCAAAAGAATTGTTCAGCACACATCTTTTGAAAGTAAGAAGAAAAACCAGATGACCA ACTATGTAATGATAACCTGTGACATCATGGACCACTCCATCTCCCCATTTATGAGGAA AGGGACCGTTGGAGAGTGGAAGGATTACTTCTCAGCAGCACAGAATAAGAGATTTGAT GAAGACAGGAAAATGGCTGACTCTTCTCTGACCTTCCACACGGAGCTCTAAAGAGAGA GAGACAAAGTCTATACTACACAGGGGCAC
	ORF Start: ATG at 34	ORF Stop: TAA at 919
	SEQ ID NO: 150	295 aa	MW at 34853.7kD
NOV48a, CG57851-01 Protein Sequence	MAKEETSELEWGLLPPEEFSQVNGIILQKKMCDFWDKIWNFQAKPDDLLIASYPKAGT TWTQEIVDLIQNDGDIEKSRRASIQLQHPFLEWIRMTHARKIFAGIDQANTMPSPRTL KTHLPVQLLPPSFWEENCKIIYVARNAKDNLVSYYHFQRMSKALPDVLTVGEYIMCGE VLWGIWEEIRTWQLHRLFCWFFDHASENPRKFKRIMEFMGNKLDEDPVKRIVQHTSFE SKKKNQMTNYVMITCDIMDHSISPFMRKGTVGEWKDYFSAAQNKRFDEDRKMADSSLT FHTEL	

Further analysis of the NOV48a protein yielded the following properties shown in Table 48B.

Table 48B. Protein Sequence Properties NOV48a	
PSort analysis:	0.6400 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

240

A search of the NOV48a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 48C.

Table 48C. Geneseq Results for NOV48a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV48a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE12209	Human ST drug-metabolising protein 2 encoded by DNA transcript 2 - Homo sapiens, 304 aa. [WO200172977-A2, 04-OCT-2001]	16..295 15..304	137/293 (46%) 200/293 (67%)	9e-74
AAE12210	Human ST drug-metabolising protein 3 encoded by cDNA - Homo sapiens, 304 aa. [WO200172977-A2, 04-OCT-2001]	16..295 15..304	129/293 (44%) 190/293 (64%)	1e-67
AAE12208	Human ST drug-metabolising protein 1 encoded by DNA transcript 1 - Homo sapiens, 304 aa. [WO200172977-A2, 04-OCT-2001]	16..295 15..304	128/293 (43%) 190/293 (64%)	6e-67
AAE05178	Human drug metabolising enzyme (DME-9) protein - Homo sapiens, 304 aa. [WO200151638-A2, 19-JUL-2001]	16..295 15..304	128/293 (43%) 189/293 (63%)	1e-66
AAY67294	Human STP2 (phenol sulphotransferase 2) amino acid sequence - Homo sapiens, 295 aa. [WO9964630-A1, 16-DEC-1999]	15..295 10..295	133/292 (45%) 186/292 (63%)	5e-66

In a BLAST search of public sequence databases, the NOV48a protein was found to have homology to the proteins shown in the BLASTP data in Table 48D.

Table 48D. Public BLASTP Results for NOV48a				
Protein Accession Number	Protein/Organism/Length	NOV48a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q90WR6	SULFOTRANSFERASE 1C - Gallus gallus (Chicken), 307 aa.	3..295 5..307	170/304 (55%) 218/304 (70%)	3e-94

241

P50237	N-hydroxyarylamine sulfotransferase (EC 2.8.2.-) (HAST-I) - Rattus norvegicus (Rat), 304 aa.	1..295 1..304	172/308 (55%) 222/308 (71%)	3e-92
O70262	PHENOL SULFOTRANSFERASE - Mus musculus (Mouse), 304 aa.	18..295 19..304	164/289 (56%) 215/289 (73%)	1e-91
O75897	Sulfotransferase 1C2 (EC 2.8.2.-) (SULT1C) (SULT1C#2) - Homo sapiens (Human), 302 aa.	22..292 22..299	160/282 (56%) 203/282 (71%)	1e-87
O00338	Sulfotransferase 1C1 (EC 2.8.2.-) (SULT1C#1) (ST1C2) (humSULTC2) - Homo sapiens (Human), 296 aa.	18..295 12..296	149/289 (51%) 201/289 (68%)	1e-80

PFam analysis predicts that the NOV48a protein contains the domains shown in the Table 48E.

Table 48E. Domain Analysis of NOV48a			
Pfam Domain	NOV48a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Sulfotransfer: domain 1 of 1	23..285	116/298 (39%) 207/298 (69%)	6.2e-82

Example 49.

The NOV49 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 49A.

Table 49A. NOV49 Sequence Analysis		
	SEQ ID NO: 151	1934 bp
NOV49a, CG59377-01 DNA Sequence	CTTGATTACGGAGACTGAACCTTCATAGGGTGCGCACTTACCAAGGACAGGAAGGTTT CTCTGTTTGAAGGGCTTTAAACTTATAACAAAGAAAATAAAAATGACGACTTCGTCTA TCAGACGGCAGATGAAAAACATCGTGAACAATTACTCAGAGGCAGAAATCAAAGTCCG GGAAGCCACCTCCAATGACCCGTGGGGCCCGTCCAGTTCTCTGATGACCGAGATTGCC GACCTGACCTACAACGTGGTGGCCTTCTCGGAGATCATGAGCATGGTGTGGAAGCGGC TGAATGACCATGGCAAGAACTGGCGGCATGTGTACAAGGCGCTGACCCTGCTGGACTA CCTCATCAAGACAGGCTCCGAACGTGTGGCCCAGCAGTGCCGGGAGAACATCTTCGCC ATCCAGACCCTGAAGGACTTCCAGTACATTGACCGAGATGGCAAGGACCAGGGCATCA ATGTGCGTGAGAAGTCAAAGCAACTGGTGGCTCTCCTCAAGGACGAGGAACGGTTGAA GGCTGAGAGGGCCCAGGCTCTCAAAACCAAAGAGCGCATGGCCCAGGTTGCCACTGGC ATGGGCAGCAACCAGATCACCTTTGGGCGAGGCTCCAGCCAGCCCAACCTCTCCACCA GCCACTCGGAGCAGGAGTATGGCAAGGCCGGGGGCTCCCCGGCCTCCTACCATGGCTC CACCTCCCCGCGAGTGTCCTCCGAGCTGGAGCAAGCCCGGCCCCAGACTAGTGGAGAA GAGGAGCTTCAGCTGCAGCTGGCACTTGCCATGAGCAGAGAAGTGGCTGAGCAGGAAG AACGCCTCAGGCGGGGTGATGACCTCAGATTACAGATGGCCCTGGAAGAAAGCCGAAG GGACACAGTTAAAATTCCAAAAAAGAAAGAGCAGACTACGCTGTTGGATTTAATGGAT GCTCTCCCCAGCTCGGGCCCCGCGGCCCAGAAAGCAGAGCCCTGGGGCCCGTCAGCCT CCACTAACCAGACCAACCCCTGGGGCGGGCCAGCGGCTCCTGCGAGTACTTCAGACCC CTGGCCATCGTTTGGTACCAAGCCAGCTGCCTCCATTGACCCATGGGGGGTGCCCACT

242

	GGAGCCACCGCACAATCTGTCCCCAAGAACTCGGACCCCTGGGCAGCTTCACAGCAGC CTGCCTCCAGTGCTGGGAAAAGAGCTTCTGACGCGTGGGGCGCAGTCTCCACCACCAA GCCCGTGTCTGTCTCTGGGTCCTTTGAGCTCTTCAGTAATCTGAATGGTACAATTAAA GATGACTTTTCTGAATTTGACAACCTTCGGACTTCAAAAAAAACAGCCGAATCTGTGA CCTCTCTGCCATCCCAAAACAATGGAACTACCAGCCCTGACCCCTTTGAGTCTCAACC CCTGACTGTCGCCTCAAGCAAGCCCAGCAGTGCCCGGAAAACACCTGAGTCCTTCCTG GGCCCCAACGCGGCCCTGGTGAACCTGGACTCACTGGTGACCAGGCCTGCCCCACCAG CCCAGTCCCTCAACCCTTTCCTGGCACCAGGTGCTCCCGCCACCTCGGCCCCTGTTAA CCCTTTCCAGGTGAACCAGCCCCAGCCGCTGACACTGAACCAGCTTCGGGGGAGCCCA GTCCTGGGGACCAGCACATCCTTTGGGCCTGGCCCAGGAGTGGAGTCCATGGCTGTGG CCTCGATGACCTCCGCGGCCCCACAGCCAGCTCTGGGGGCCACTGGTTCCTCTCTGAC ACCACTGGGCCCTGCAATGATGAACATGGTGGGCAGTGTGGGTATACCCCCATCAGCA GCCCAGGCCACTGGCACAACCAACCCTTTCCTTCTCTAGTGCCTGGGCCTGGGACCCA CCCAGAGCACCTGTGCTGGAGGATGCCGAGCAGGGACTCTCGTCTGTGGGACGGGATC CAAGAGTTTGGGGATTAGGG
	ORF Start: ATG at 101	ORF Stop: TAG at 1835
	SEQ ID NO: 152	578 aa	MW at 61651.2kD
NOV49a, CG59377-01 Protein Sequence	MTTSSIRRQMKNIVNNYSEAEIKVREATSNDPWGPSSSLMTEIADLTYNVVAFSEIMS MVWKRLNDHGKNWRHVYKALTLLDYLIKTGSERVAQQCRENIFAIQTLKDFQYIDRDG KDQGINVREKSKQLVALLKDEERLKAERAQALKTKERMAQVATGMGSNQITFGRGSSQ PNLSTSHSEQEYGKAGGSPASYHGSTSPRVSSELEQARPQTSGEEELQLQLALAMSRE VAEQEERLRRGDDLRLQMALEESRRDTVKIPKKKEQTTLLDLMDALPSSGPAAQKAEP WGPSASTNQTNPWGGPAAPASTSDPWPSFGTKPAASIDPWGVPTGATAQSVPKNSDPW AASQQPASSAGKRASDAWGAVSTTKPVSVSGSFELFSNLNGTIKDDFSEFDNLRTSKK TAESVTSLPSQNNGTTSPDPFESQPLTVASSKPSSARKTPESFLGPNAALVNLDSLVT RPAPPAQSLNPFLAPGAPATSAPVNPFQVNQPQPLTLNQLRGSPVLGTSTSFGPGPGV ESMAVASMTSAAPQPALGATGSSLTPLGPAMMNMVGSVGIPPSAAQATGTTNPFLL

Further analysis of the NOV49a protein yielded the following properties shown in Table 49B.

Table 49B. Protein Sequence Properties NOV49a	
PSort analysis:	0.4936 probability located in mitochondrial matrix space; 0.3000 probability located in nucleus; 0.2087 probability located in mitochondrial inner membrane; 0.2087 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV49a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 49C.

Table 49C. Geneseq Results for NOV49a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB93525	Human protein sequence SEQ ID NO:12872 - Homo sapiens, 584 aa. [EP1074617-A2, 07-FEB-2001]	1..578 1..584	578/584 (98%) 578/584 (98%)	0.0

AAB95663	Human protein sequence SEQ ID NO:18438 - Homo sapiens, 370 aa. [EP1074617-A2, 07-FEB-2001]	40..403 1..370	364/370 (98%) 364/370 (98%)	0.0
AAB93011	Human protein sequence SEQ ID NO:11762 - Homo sapiens, 484 aa. [EP1074617-A2, 07-FEB-2001]	1..407 1..470	385/470 (81%) 390/470 (82%)	0.0
AAB42049	Human ORFX ORF1813 polypeptide sequence SEQ ID NO:3626 - Homo sapiens, 551 aa. [WO200058473-A2, 05-OCT-2000]	1..578 1..551	306/636 (48%) 370/636 (58%)	e-141
AAB95100	Human protein sequence SEQ ID NO:17064 - Homo sapiens, 576 aa. [EP1074617-A2, 07-FEB-2001]	1..578 1..576	298/636 (46%) 371/636 (57%)	e-137

In a BLAST search of public sequence databases, the NOV49a protein was found to have homology to the proteins shown in the BLASTP data in Table 49D.

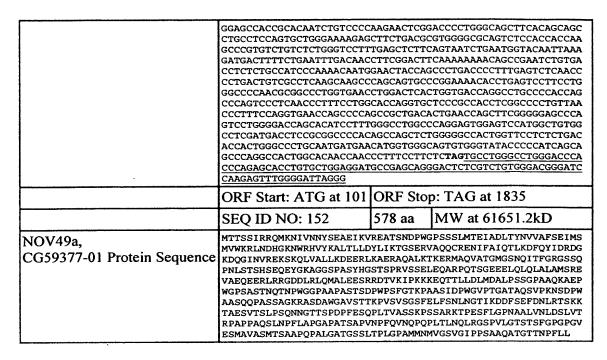| Table 49D. Public BLASTP Results for NOV49a				
Protein Accession Number	Protein/Organism/Length	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O95207	EPSIN 2A - Homo sapiens (Human), 584 aa.	1..578 1..584	576/584 (98%) 576/584 (98%)	0.0
Q9UPT7	KIAA1065 PROTEIN - Homo sapiens (Human), 641 aa.	1..578 1..641	557/641 (86%) 562/641 (86%)	0.0
O95208	EPSIN 2B - Homo sapiens (Human), 642 aa.	1..578 1..642	556/642 (86%) 560/642 (86%)	0.0
Q9Z1Z3	EH DOMAIN BINDING PROTEIN EPSIN 2 - Rattus norvegicus (Rat), 583 aa.	1..578 1..583	512/590 (86%) 526/590 (88%)	0.0
O70447	INTERSECTIN-EH BINDING PROTEIN IBP2 - Mus musculus (Mouse), 509 aa (fragment).	76..578 2..509	438/515 (85%) 459/515 (89%)	0.0

PFam analysis predicts that the NOV49a protein contains the domains shown in the Table 49E.

244

Table 49E. Domain Analysis of NOV49a			
Pfam Domain	NOV49a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ENTH: domain 1 of 1	17..140	70/131 (53%) 117/131 (89%)	7.9e-68
VHS: domain 1 of 1	14..158	33/160 (21%) 90/160 (56%)	3.3
UIM: domain 1 of 2	217..234	11/18 (61%) 16/18 (89%)	0.043
UIM: domain 2 of 2	242..259	5/18 (28%) 12/18 (67%)	80

Example 50.

The NOV50 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 50A.

Table 50A. NOV50 Sequence Analysis		
	SEQ ID NO: 153	2580 bp
NOV50a, CG59258-01 DNA Sequence	ATGCTGCTGGCCCCCTTTTATTGCTGGGTGTGTGCCCATGCTGCTGGCCCCCTTTTAT TGCTGGGCAGTGACAAACTGTACCATCAGTGGCTCTCCACTGTCCGGAAAGGAAGTGG AGCAATTCTGAATACTGTAAAGACCAAAGCAAATCCGGCCATGAAGACTGTCTACAAG TTCGACATTGCCGAGAATGGCTGCGCCCCCACCCCAGAAGAGCAGCTGCCAAAGACTG CACCGTCCCCACTGGTGGAGGCCAAGGACCCCAAGCTCCGAGAAGACCGGCGGCCAAT CACAGTCCACTTTGGACAGGACCAGTCTGAGATGTCTTTCAGCTCAGCACTCACTCAC GGCAAAGAGAGTGCCCGGACCCAGCCGGAGAGAGTCGTTGACAGGACTGGCGAGCCCC TGAATCCTGAGCGCGCTCTCTCCGGAGATCATCTCTGGCCTGTTACGCACTTGCTCTG GGCAACCCTGGGCAAGTCCTTGCTTGCCCTCATCTGTGAAATGGGTAGCAGCCCTCGT TCCCTGCAGAGGAGCCTTGCGCTGCTGGGGACACCCCAGCTTATTTGGGAAACTGCAA CCACCATGGCCGATGGCCCCACCACGCCCTGTCTAGGAAGCAGAGGCCTCCCCAGCAG CGTGTCCACTGTGCCCCTGGCCCTGCGTGAAGTGCCATCAGATGCCCCGCATCCCTGC AGCAGGGCCCTCGTGACTGGCCTCACAGATGAGGACACAGAGGCCCAGGGAAGTCACT TGCTTGCCAAAGTCACTCAGCAAACCATGTCTGTCTGGCTCTCAGAAAATGGGAAAGA AGCCTGGGCATTCAGCCATGAGGGAGCCACGGCTGTAGCCAGTGGAATGACGTACCCT CAGTCCAGGATGTGCACCCGGGCAGCCAGGTCCCACAGCCACTACTTTCTTGCCCCCA CCACTGCTCCCACAGTTCCCAGAACTCAGTCTCCAGATCTGGGCTCCAGGATGCAGAG GCTGTCCTCAGGCCTGGTAAAGCCCTTGCGACACTATGCGGTCTTCCTCTCCGAAGAC TCCTCTGATGATGAATGCCAGCGGGAAGAGGGCCCGAGCTCTGGCTTCACCGAGAGCT TTTTCTTCTCCGCTCCCTTTGAATGGCCGCAGCCGTATCGGACACTCAGGGAGTCAGA CAGCGCGGAAGGCGACGAGGCAGAGAGTCCAGAGCAGCAAGTGCGGAAGTCCACAGGC CCTGTCCCAGCTCCCCCTGACCGGGCTGCCAGCATCGACCTTCTGGAAGACGTCTTCA GCAACCTGGACATGGAGGCCGCACTGCAGCCCACTGGGCCAGGCCAAGAGCTTAGAGGA CCTTCGTGCCCCCAAAGACCTGAGGGAGCAGCCAGGGACCTTTGACTATCAGAGGCTG GATCTGGGCGGGAGTGAGAGGAGCCGCGGGGTGACAGTGGCCTTGAAGCTTACCCACC CGTACAACAAGCTCTGGAGCCTGGGCCAGGACGACATGGCCATCCCCAGCAAGCCCCC AGCTGCCTCCCCTGAGAAGCCCTCAGCCCTGCTCGGAAACTCCCTGGCCCTGCCTCGA AGGCCCCAGAACCGGGACAGCATCCTGAACCCCCAGTGACAAGGAGGAGGTGCCCACCC CTACTCTGGGCAGCATCACCATCCCCCGGCCCCAAGGCAGGAAGACCCCAGAGCTGGG CATCGTGCCTCCACCGCCCATTCCCCGCCCGGCCAAGCTCCAGGCTGCCGGCGCCGCA CTTGGTGACGTCTCAGAGCGGCTGCAGACGGATCGGGACAGGCGAGCTGCCCTGAGTC CAGGGCTCCTGCCTGGTGTTGTCCCCCAAGGCCCCACTGAACTGCTCCAGCCGCTCAG CCCTGGCCCCGGGGCTGCAGGCACGAGCAGTGACGCCCTGCTCGCCCTCCTGGACCCG CTCAGCACAGCCTGGTCAGGCAGCACCCTCCCGTCACGCCCCGCCACCCCGAATGTAG CCACCCCATTCACCCCCCAATTCAGCTTCCCCCCTGCAGGGACACCCACCCCATTCCC ACAGCCACCACTCAACCCCTTTGTCCCATCCATGCCAGCAGCCCCACCCACCCTGCCC CTGGTCTCCACACCAGCCGGGCCTTTCGGGGCCCCTCCAGCTTCCCTGGGGCCGGCTT TTGCGTCCGGCCTCCTGCTGTCCAGTGCTGGCTTCTGTGCCCCTCACAGGTCTCAGCC CAACCTCTCCGCCCTCTCCATGCCCAACCTCTTTGGCCAGATGCCCATGGGCACCCAC ACGAGCCCCCTACAGCCGCTGGGTCCCCCAGCAGTTGCCCCGTCGAGGATCCGAACGT

245

	TGCCCCTGGCCCGCTCAAGTGCCAGGGCTGCTGAGACCAAGCAGGGGCTGGCCCTGAG GCCTGGAGACCCCCCGCTTCTGCCTCCCAGGCCCCCTCAAGGCCTGGAGCCAACACTG CAGCCCTCTGCTCCTCAACAGGCCAGAGACCCCTTTGAGGATTTGTTACAGAAAACCA AGCAAGACGTGAGCCCGAGTCCGGCCCTGGCCCCGGCCCCAGACTCGGTGGAGCAGCT CAGGAAGCAGTGGGAGACCTTCGAGTGA
	ORF Start: ATG at 1 | ORF Stop: TGA at 2578
	SEQ ID NO: 154 | 859 aa | MW at 91746.7kD
NOV50a, CG59258-01 Protein Sequence	MLLAPFYCWVCAHAAGPLLLLGSDKLYHQWLSTVRKGSGAILNTVKTKANPAMKTVYK FDIAENGCAPTPEEQLPKTAPSPLVEAKDPKLREDRRPITVHFGQDQSEMSFSSALTH GKESARTQPERVVDRTGEPLNPERALSGDHLWPVTHLLWATLGKSLLALICEMGSSPR SLQRSLALLGTPQLIWETATTMADGPTTPCLGSRGLPSSVSTVPLALREVPSDAPHPC SRALVTGLTDEDTEAQGSHLLAKVTQQTMSVWLSENGKEAWAFSHEGATAVASGMTYP QSRMCTRAARSHSHYFLAPTTAPTVPRTQSPDLGSRMQRLSSGLVKPLRHYAVFLSED SSDDECQREEGPSSGFTESFFFSAPFEWPQPYRTLRESDSAEGDEAESPEQQVRKSTG PVPAPPDRAASIDLLEDVFSNLDMEAALQPLGQAKSLEDLRAPKDLREQPGTFDYQRL DLGGSERSRGVTVALKLTHPYNKLWSLGQDDMAIPSKPPAASPEKPSALLGNSLALPR RPQNRDSILNPSDKEEVPTPTLGSITIPRPQGRKTPELGIVPPPPIPRPAKLQAAGAA LGDVSERLQTDRDRRAALSPGLLPGVVPQGPTELLQPLSPGPGAAGTSSDALLALLDP LSTAWSGSTLPSRPATPNVATPFTPQFSFPPAGTPTPFPQPPLNPFVPSMPAAPPTLP LVSTPAGPFGAPPASLGPAFASGLLLSSAGFCAPHRSQPNLSALSMPNLFGQMPMGTH TSPLQPLGPPAVAPSRIRTLPLARSSARAAETKQGLALRPGDPPLLPPRPPQGLEPTL QPSAPQQARDPFEDLLQKTKQDVSPSPALAPAPDSVEQLRKQWETFE

Further analysis of the NOV50a protein yielded the following properties shown in Table 50B.

Table 50B. Protein Sequence Properties NOV50a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1940 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	Likely cleavage site between residues 15 and 16

A search of the NOV50a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 50C.

Table 50C. Geneseq Results for NOV50a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV50a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41501	Human polypeptide SEQ ID NO 6432 - Homo sapiens, 545 aa. [WO200153312-A1, 26-JUL-2001]	22..103 401..482	82/82 (100%) 82/82 (100%)	2e-42
AAM39715	Human polypeptide SEQ ID NO 2860 - Homo sapiens, 559 aa. [WO200153312-A1, 26-JUL-2001]	22..103 396..496	82/101 (81%) 82/101 (81%)	6e-39

AAW31855	Mycobacterium tuberculosis 55 kDa protein - Mycobacterium tuberculosis, 572 aa. [WO9741252-A2, 06-NOV-1997]	498..845 71..389	96/358 (26%) 125/358 (34%)	8e-12
AAW31852	Mycobacterium tuberculosis 74 kDa protein - Mycobacterium tuberculosis, 763 aa. [WO9741252-A2, 06-NOV-1997]	498..845 262..580	96/358 (26%) 125/358 (34%)	8e-12
AAB50363	Human SRCAP - Homo sapiens, 2972 aa. [WO200073467-A1, 07-DEC-2000]	501..845 1235..1575	112/369 (30%) 141/369 (37%)	1e-11

In a BLAST search of public sequence databases, the NOV50a protein was found to have homology to the proteins shown in the BLASTP data in Table 50D.

Table 50D. Public BLASTP Results for NOV50a				
Protein Accession Number	Protein/Organism/Length	NOV50a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9HCG4	KIAA1608 PROTEIN - Homo sapiens (Human), 603 aa (fragment).	309..859 62..603	501/555 (90%) 510/555 (91%)	0.0
Q9H796	CDNA: FLJ21129 FIS, CLONE CAS06266 - Homo sapiens (Human), 559 aa.	22..103 396..496	81/101 (80%) 81/101 (80%)	2e-37
AAK44515	HYPOTHETICAL 58.5 KDA PROTEIN - Mycobacterium tuberculosis CDC1551, 598 aa.	499..845 299..562	104/354 (29%) 121/354 (33%)	8e-14
Q9SN46	EXTENSIN-LIKE PROTEIN - Arabidopsis thaliana (Mouse-ear cress), 839 aa.	604..848 407..626	73/249 (29%) 100/249 (39%)	3e-12
Q41805	EXTENSIN-LIKE PROTEIN PRECURSOR - Zea mays (Maize), 1188 aa.	492..848 415..749	88/361 (24%) 124/361 (33%)	5e-12

PFam analysis predicts that the NOV50a protein contains the domains shown in the Table 50E.

5

247

Table 50E. Domain Analysis of NOV50a			
Pfam Domain	NOV50a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 51.

The NOV51 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 51A.

Table 51A. NOV51 Sequence Analysis	
	SEQ ID NO: 155 1394 bp
NOV51a, CG59492-01 DNA Sequence	GTGGCCTGCTCCTGCAGCAATCCCAGGACCCCCTGCTCATGGGGCTGTTTCCTACTAA CCCCAAAGAGAAGACCCAGGAGGAACCCCCTGGCCAGAGCAGGGCCCCTGTGTTGACC GTGGTGTCCAAGTTCAAGGCCTCACTGGAGCAGCTTCTGCAGGTCCTACACAGCACCA CGCCCCACTACATTCGCTGCATCAAGCCCAACAGCCAGGGCCAGGCGCAGACCTTTCT CCAAGAGGAGGTCCTGAGCCAGCTGGAGGCCTGTGGCCTCGTGGAGACCATCCATATC AGTGCTGCTGGCTTCCCCATCCGGGTCTCTCACCGAAACTTTGTAGAACGATACAAGT TACTAAGAAGGCTTCATCCTTGCACATCCTCTGGCCCCGACAGCCCATATCCTGCCAA AGGGCTCCCTGAATGGTGTCCACACAGCGAGGAAGCCACGCTTGAACCTCTCATCCAG GACATTCTCCACACTCTGCCGGTCCTAACTCAGGCAGCAGCCATAACTGGTGACTCGG CTGAGGCCATGCCAGCCCCCATGCACTGTGGCAGGACCAAGGTGTTCATGACTGACTC TATGCTGGAGCTTCTGGAATGTGGCGTGCCCGGGTGCTGGAGCAGTGTGCCCGCTGC ATCCAGGGTGGCTGGAGGCGACACCGGCACCGAGAGCAGGAGCGGCAGTGGCGGGCCG TCATGCTCATCCAGGCAGCCATTCGTTCCTGGTTAACTCGGAAACACATCCAGAGGCT GCATGCAGCTGCCACAGTCATCAAGCGTGCATGGCAGAAGTGGAGAATCAGAATGGCC TGCCTTGCTGCTAAAGAGCTGGATGGTGTGGAAGAAAAACACTTCTCTCAAGCTCCCT GTTCCCTGAGCACCTCGCCGCTGCAGACCAGGCTCCTGGAGGCAATAATCCGCTTCTG GCCCCTGGGACTGGTCCTGGCCAATACGGCTATGGGTGTAGGCAGCTTTCAGAGGAAA TTAGTGGTCTGGGCTTGCCTCCAGCTCCCCAGGGGCAGCCCCAGTAGCTACACTGTCC AGACAGCACAAGACCAGGCTGGTGTCACGTCCATCCGAGCGCTGCCTCAGGGATCGAT AAAGTTTCACTGCAGAAAGTCTCCACTGCGGTATGCTGACATCTGCCCTGAACCTTCA CCCTACAGCATTGCAGGCTTTAATCAGATTCTGCTGGAAAGACACAGGCTGATCCACG TGACCTCTTCTGCCTTCACTGGGCTGGGGTGATCCTTGGTGCCTTTGTTTCCACAAGG CCTTTTCCTGCCCCCTGCCTTGCCAAAGACATTTAATCAGCACACAGCTGCCAGACTA TTCCCACAGTGCTCCAAATGCACATGAACAACAGTGACGGCTCCAGCCTTCGACCCAG AG
	ORF Start: ATG at 39 ORF Stop: TGA at 1248
	SEQ ID NO: 156 403 aa MW at 45142.8kD
NOV51a, CG59492-01 Protein Sequence	MGLFPTNPKEKTQEEPPGQSRAPVLTVVSKFKASLEQLLQVLHSTTPHYIRCIKPNSQ GQAQTFLQEEVLSQLEACGLVETIHISAAGFPIRVSHRNFVERYKLLRRLHPCTSSGP DSPYPAKGLPEWCPHSEEATLEPLIQDILHTLPVLTQAAAITGDSAEAMPAPMHCGRT KVFMTDSMLELLECGRARVLEQCARCIQGGWRRHRHREQERQWRAVMLIQAAIRSWLT RKHIQRLHAAATVIKRAWQKWRIRMACLAAKELDGVEEKHFSQAPCSLSTSPLQTRLL EAIIRFWPLGLVLANTAMGVGSFQRKLVVWACLQLPRGSPSSYTVQTAQDQAGVTSIR ALPQGSIKFHCRKSPLRYADICPEPSPYSIAGFNQILLERHRLIHVTSSAFTGLG

Further analysis of the NOV51a protein yielded the following properties shown in Table 51B.

Table 51B. Protein Sequence Properties NOV51a	
PSort analysis:	0.3000 probability located in nucleus; 0.2029 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space; 0.0320 probability located in microbody (peroxisome)

248

SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV51a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 51C.
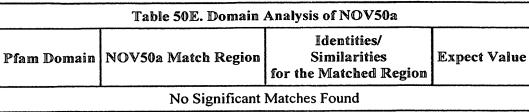
Table 51C. Geneseq Results for NOV51a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV51a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY94290	Human myosin heavy chain homologue - Homo sapiens, 612 aa. [WO200026372-A1, 11-MAY-2000]	1..403 210..612	401/403 (99%) 401/403 (99%)	0.0
AAU23676	Novel human enzyme polypeptide #762 - Homo sapiens, 387 aa. [WO200155301-A2, 02-AUG-2001]	17..403 1..387	384/387 (99%) 384/387 (99%)	0.0
ABB10243	Human cDNA SEQ ID NO: 551 - Homo sapiens, 570 aa. [WO200154474-A2, 02-AUG-2001]	1..365 206..570	365/365 (100%) 365/365 (100%)	0.0
AAU23123	Novel human enzyme polypeptide #209 - Homo sapiens, 567 aa. [WO200155301-A2, 02-AUG-2001]	1..365 203..567	364/365 (99%) 364/365 (99%)	0.0
AAM23563	Human EST encoded protein SEQ ID NO: 1088 - Homo sapiens, 477 aa. [WO200154477-A2, 02-AUG-2001]	1..189 288..476	188/189 (99%) 188/189 (99%)	e-108

In a BLAST search of public sequence databases, the NOV51a protein was found to have homology to the proteins shown in the BLASTP data in Table 51D.

Table 51D. Public BLASTP Results for NOV51a				
Protein Accession Number	Protein/Organism/Length	NOV51a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96H55	HYPOTHETICAL 86.7 KDA PROTEIN - Homo sapiens (Human), 770 aa.	72..403 439..770	330/332 (99%) 330/332 (99%)	0.0
Q9D2Z3	1110055A02RIK PROTEIN (RIKEN CDNA 1110055A02 GENE) - Mus musculus (Mouse), 395 aa.	3..394 2..395	288/394 (73%) 320/394 (81%)	e-162
Q948A2	PUTATIVE MYOSIN HEAVY CHAIN - Oryza sativa (Rice), 1601 aa.	2..255 663..876	84/258 (32%) 125/258 (47%)	1e-23
O74805	HYPOTHETICAL MYOSIN-LIKE PROTEIN C2D10.14C IN CHROMOSOME II - Schizosaccharomyces pombe (Fission yeast), 1471 aa.	20..347 615..903	96/340 (28%) 152/340 (44%)	1e-21
T30148	hypothetical protein E02C12.1 - Caenorhabditis elegans, 1019 aa.	5..249 619..830	74/248 (29%) 119/248 (47%)	6e-21

PFam analysis predicts that the NOV51a protein contains the domains shown in the Table 51E.

Table 51E. Domain Analysis of NOV51a			
Pfam Domain	NOV51a Match Region	Identities/ Similarities for the Matched Region	Expect Value
myosin_head: domain 1 of 1	26..105	37/81 (46%) 60/81 (74%)	5.1e-25

5

Example 52.

The NOV52 clone was analyzed, and the nucleotide and predicted polypeptide

10   sequences are shown in Table 52A.

Table 52A. NOV52 Sequence Analysis		
	SEQ ID NO: 157	1380 bp
NOV52a, CG59564-01 DNA Sequence	TAGAATTCCAGCGGCCGCTGAAATCCTCACTCGGTCAGTTCCTCGGGCGAGTTACGGG GACGACCTGCGGGAGCACGCGGGCAGTGGCCGGACGCTGAAGCCCAGGAGAGCGATGG AGACGTATGCGGAGGTTGGGAAGGAGGGCAAGCCTTCCTGTGCATCGGTGGATCTGCA GGGAGACAGCTCCTTACAGGTGGAGATTTCTGACGCAGTGAGTGAGCGGGACAAGGTG AAATTCACTGTTCAAACAAAGAGCTGCCTCCCTCACTTCGCCCAGACCGAGTTCTCAG TCGTGCGGCAGCACGAGGAGTTCATCTGGCTGCATGATGCCTACGTGGAGAATGAGGA GTACGCCGGCCTCATCATCCCCCCAGCCCCTCCGAGGCCAGACTTTGAGGCTTCGAGG GAAAAGCTACAGAAATTGGGCGAGGGGGACAGCTCTGTCACTCGGGAAGAGTTTGCCA AGATGAAGCAGGAGCTGGAAGCGGAGTACCTGGCCATCTTTAAGAAGACAGTTGCGAT GCACGAAGTCTTTCTGCAGCGCCTGGCGGCCCACCCCACCCTGCGTCGAGACCACAAC TTCTTTGTGTTTTTGGAATATGGACAGGATCTGAGTGTCCGGGGGAAGAACAGGAAGG AGCTCCTCGGAGGGTTTCTGAGGAATATTGTGAAGTCCGCGGATGAAGCCCTCATCAC GGGCATGTCAGGGCTCAAGGAGGTGGATGACTTCTTTGAGCATGAGAGGACCTTCCTG TTGGAGTATCACACCCGTATCCGAGATGCCTGCCTGCGGGCCGACCGCGTCATGCGCG CCCACAAGTGCCTGGCAGACGATTATATCCCTATCTCAGCTGCGCTGAGCAGTCTGGG AACACAGGAAGTCAACCAGCTAAGGACGAGCTTCCTCAAATTGGCAGAGCTCTTTGAC CGGCTGAGGAAGCTGGAGGGCCGGGTGGCTTCCGATGAGGACCTGAAGCTGTCAGACA TGCTGAGGTACTACATGCGTGACTCACAGGCAGCCAAGGACCTGCTGTACCGGCGGCT GCGGGCACTGGCCGACTACGAGAATGCCAACAAGGCGCTGGACAAGGCGCGCACCAGG AACCGGGAGGTGCGGCCCGCCGAGAGCCACCAGCAGCTGTGCTGCCAACGCTTCGAGC GCCTCTCCGACTCCGCCAAGCAAGAGCTCATGGACTTCAAGTCCCGCCGGGTCTCCTC TTTTCGAAAGAATCTCATTGAGCTGGCAGAGCTGGAGCTCAAACACGCCAAGGCCAGC ACCCTGATTCTCCGGAACACCCTTGTTGCCCTAAAGGGGGAGCCTTAGAGTAGCCAGA GCTCAGCCAGACCCTAATCTGGGATCTCCAGTGACCAGGGTATCCC
	ORF Start: ATG at 113	ORF Stop: TAG at 1322
	SEQ ID NO: 158	403 aa	MW at 46384.2kD
NOV52a, CG59564-01 Protein Sequence	METYAEVGKEGKPSCASVDLQGDSSLQVEISDAVSERDKVKFTVQTKSCLPHFAQTEF SVVRQHEEFIWLHDAYVENEEYAGLIIPPAPPRPDFEASREKLQKLGEGDSSVTREEF AKMKQELEAEYLAIFKKTVAMHEVFLQRLAAHPTLRRDHNFFVFLEYGQDLSVRGKNR KELLGGFLRNIVKSADEALITGMSGLKEVDDFFEHERTFLLEYHTRIRDACLRADRVM RAHKCLADDYIPISAALSSLGTQEVNQLRTSFLKLAELFDRLRKLEGRVASDEDLKLS DMLRYYMRDSQAAKDLLYRRLRALADYENANKALDKARTRNREVRPAESHQQLCCQRF ERLSDSAKQELMDFKSRRVSSFRKNLIELAELELKHAKASTLILRNTLVALKGEP	

Further analysis of the NOV52a protein yielded the following properties shown in Table 52B.

Table 52B. Protein Sequence Properties NOV52a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV52a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 52C.

251

Table 52C. Geneseq Results for NOV52a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV52a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY94209	Human TRAF four associated factor TFAF2 - Homo sapiens, 406 aa. [CA2245340-A1, 19-FEB-2000]	17..402 23..405	273/386 (70%) 333/386 (85%)	e-160
AAB07856	Amino acid sequence of Smad1 interactor protein clone S1+12-2 - Homo sapiens, 414 aa. [WO200047102-A2, 17-AUG-2000]	17..402 31..413	273/386 (70%) 333/386 (85%)	e-160
AAB43157	Human ORFX ORF2921 polypeptide sequence SEQ ID NO:5842 - Homo sapiens, 460 aa. [WO200058473-A2, 05-OCT-2000]	17..402 77..459	273/386 (70%) 333/386 (85%)	e-160
AAB58368	Lung cancer associated polypeptide sequence SEQ ID 706 - Homo sapiens, 414 aa. [WO200055180-A2, 21-SEP-2000]	17..402 31..413	273/386 (70%) 333/386 (85%)	e-160
AAO13507	Human polypeptide SEQ ID NO 27399 - Homo sapiens, 443 aa. [WO200164835-A2, 07-SEP-2001]	17..400 61..441	242/384 (63%) 317/384 (82%)	e-144

In a BLAST search of public sequence databases, the NOV52a protein was found to have homology to the proteins shown in the BLASTP data in Table 52D.

Table 52D. Public BLASTP Results for NOV52a				
Protein Accession Number	Protein/Organism/Length	NOV52a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UNH7	Sorting nexin 6 (TRAF4-associated factor 2) - Homo sapiens (Human), 406 aa.	17..402 23..405	273/386 (70%) 333/386 (85%)	e-159
Q9CZ03	2810425K19RIK PROTEIN - Mus musculus (Mouse), 406 aa.	17..402 23..405	271/386 (70%) 333/386 (86%)	e-159
Q9Y5X3	Sorting nexin 5 - Homo sapiens (Human), 404 aa.	17..400 22..402	242/384 (63%) 317/384 (82%)	e-143

252

Q9D8U8	Sorting nexin 5 - Mus musculus (Mouse), 404 aa.	17..400 22..402	241/384 (62%) 314/384 (81%)	e-142
Q96NG4	CDNA FLJ30934 FIS, CLONE FEBRA2007017, MODERATELY SIMILAR TO HOMO SAPIENS TRAF4-ASSOCIATED FACTOR 2 MRNA - Homo sapiens (Human), 277 aa.	1..237 1..237	236/237 (99%) 236/237 (99%)	e-134

PFam analysis predicts that the NOV52a protein contains the domains shown in the Table 52E.

Table 52E. Domain Analysis of NOV52a			
Pfam Domain	NOV52a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PX: domain 1 of 1	23..164	39/160 (24%) 103/160 (64%)	1.6e-15

Example 53.

The NOV53 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 53A.

Table 53A. NOV53 Sequence Analysis		
	SEQ ID NO: 159	3056 bp
NOV53a, CG59553-01 DNA Sequence	CTCCTGCGGGGTCAAATACAGAATTTACGCACCCTTGGCTTCCTTGGAGCCTAGCGGC TCTCCCCGCGTCCAAGATGGCGGCAGAAGCAGCTGGTGGGAAATACAGAAGCACAGTC AGCAAAAGCAAAGACCCCTCGGGGCTGCTCATCTCTGTGATCAGGACTCTGTCTACTA GTGACGATGTCGAAGACAGGGAAAATGAAAAGGGTCGCCTTGAAGAAGCCTACGAGAA ATGTGACCGTGACCTGGATGAATTGATTGTACAGCACTACACAGAATTGACGACAGCC ATTCGCACATACCAGAGCATCACAGAGCGCATCACTAACTCCCGAAATAAAATAAAGC AGGTAAAAGAGAACCTGCTTTCATGCAAGATGCTGCTGCACTGCAAACGGGATGAGCT TCGGAAACTGTGGATTGAAGGAATTGAGCATAAGCATGTCCTGAACTTGTTGGATGAA ATTGAGAATATCAAGCAAGTGCCTCAAAAGCTGGAACAGTGCATGGCCAGCAAGCACT ATCTCAGTGCCACTGACATGTTGGTGTCAGCAGTTGAGTCTTTGGAGGGCCCCCTGCT CCAGGTGGAAGGACTGAGTGACCTTCGACTAGAGCTTCACAGCAAGAAGATGAACCTT CACTTGGTTCTCATAGATGAACTACACCGGCACCTGTACATCAAATCGACTAGCCGAG TTGTGCAGCGTAACAAGGAAAAAGGGAAAATCAGCTCCCTCGTGAAAGATGCTTCTGT TCCTCTGATTGATGTTACAAACCTCCCTACTCCTCGAAAATTCCTTGATACCTCTCAC TATTCTACTGCTGGAAGCTCAAGTGTGAGGGAGATAAATCTGCAGGACATCAAGGAAG ATTTAGAATTGGATCCAGAGGAAAACAGCACCCTGTTTATGGGTATCCTCATTAAGGG CTTGGCGAAACTGAAGAAGATCCCAGAAACAGTTAAGGCAATCATAGAGCGCTTGGAG CAGGAGTTGAAGCAAATTGTGAAGAGGTCTACAACCCAGGTGGCAGACAGTGGCTATC AGCGGGGGGAGAACGTTACTGTGGAGAACCAACCAAGGTTGCTTCTAGAACTGCTGGA GTTACTGTTTGACAAGTTTAATGCTGTAGCCGCTGCACACTCTGTGGTCCTGGGATAC CTGCAGGACACTGTAGTGACTCCACTGACTCAGCAGGAAGATATCAAACTGTATGATA TGGCAGATGTATGGGTGAAGATCCAAGATGTTCTACAGATGCTATTAACTGAGTACTT GGATATGAAAAATACTCGTACGGCCTCTGAACCATCAGCTCAACTAAGCTATGCCAGC ACTGGACGAGAGTTTGCAGCCTTTTTTGCCAAGAAGAAACCTCAAAGGCCAAAAAATT CTCTTTTCAAGTTCGAATCGTCCTCCCATGCCATCAGTATGAGCGCCTATCTGCGAGA ACAGAGAAGGGAGCTCTATAGTCGGAGTGGAGAACTGCAAGGGGGTCCTGATGACAAC TTAATTGAAGGTGGAGGAACAAAATTTGTCTGCAAACCTGGAGCCAGAAACATTACCG TCATATTCCACCCATTACTAAGATTTATTCAGGAGATTGAGCATGCTCTGGGTCTTGG CCCAGCCAAACAGTGTCCTCTTCGAGAGTTTCTCACCGTGTACATCAAAAACATCTTT CTCAATCAAGTCTTGGCTGAGATCAACAAGGAGATTGAAGGAGTCACTAAAACATCTG

253

	ACCCTTTGAAGATTCTGGCCAACGCAGACACCATGAAGGTGCTGGGAGTGCAGCGGCC TCTCCTACAGAGCACAATCATTGTGGAGAAGACAGTTCAAGACCTCCTGAACCTGATG CATGACTTGAGTGCATATTCAGATCAATTCCTCAACATGGTGTGCGTGAAGCTCCAGG AGTACAAGGACACCTGCACTGCAGCTTACAGGGGTATTGTCCAGTCAGAAGAAAAACT TGTCATCAGTGCATCCTGGGCAAAAGATGATGATATCAGCAGACTCTTGAAATCTCTA CCAAACTGGATGAATATGGCTCAACCCAAACAGCTGAGGCCAAAAGAGAGGAGGAAG AAGATTTCATAAGGGCAGCTTTTGGCAAGGAGTCTGAAGTTCTTATTGGGAACCTGGG TGATAAATTAATCCCTCCACAAGACATCCTTCGTGACGTCAGTGACCTCAAAGCCTTG GCCAACATGCATGAAAGCCTGGAATGGTTGGCAAGTCGAACAAAGTCAGCTTTCTCCA ATCTTTCTACATCCCAGATGCTTTCTCCTGCTCAAGCACAGCCACACGAACACGGATCT CCCCCCAGTGTCAGAGCAGATCATGCAGACTCTCAGTGAACTTGCCAAATCGTTCCAG GATATGGCTGACCGCTGCTTGCTTGTCTTACATCTGGAAGTGAGGGTTCACTGTTTCC ACTATCTTATCCCTCTTGCAAAGGAGGGGAACTATGCCATTGTGGCTAATGTGGAAAG TATGGATTATGACCCCCTGGTGGTCAAGCTCAACAAAGATATCAGCGCCATTGAAGAG GCCATGAGCGCCAGCCTTCAGCAGCACAAGTTCCAGTATATCTTCGAAGGCCTGGGCC ACCTGATCTCCTGCATCCTCATTAATGGTGCCCAGTACTTCAGGCGCATCAGTGAGTC TGGCATCAAGAAAATGTGTAGGAACATTTTTGTTCTTCAGCAGAATTTGACCAACATC ACCATGTCGCGGGAGGCAGACCTGGACTTTGCAAGGCAGTACTACGAGATGCTTTACA ACACAGCTGACGAGCTCCTGAACCTGGTGGTGGACCAGGGTGTGAAGTACACGGAGCT GGAGTACATCCACGCTCTGACCCTGCTGCACCGCAGCCAGACTGGGGTGGGGGAACTG ACCACCCAGAACACGAGCTGCAGAGGAGGCTCAAAGAGATCATCTGCGAGCAGGCTGC CATCAAGCAAGCCACCAAGGACAAGAAGATAACTACCGTTTAGCAGGGCGTACTGCGG TTGGTGACGGGGGTCCCCTCAGTCACACTCACTTTTTTCC
	ORF Start: ATG at 75 **ORF Stop: TAA at 2988**
	SEQ ID NO: 160 | **971 aa** | **MW at 109984.9kD**
NOV53a, CG59553-01 Protein Sequence	MAAEAAGGKYRSTVSKSKDPSGLLISVIRTLSTSDDVEDRENEKGRLEEAYEKCDRDL DELIVQHYTELTTAIRTYQSITERITNSRNKIKQVKENLLSCKMLLHCKRDELRKLWI EGIEHKHVLNLLDEIENIKQVPQKLEQCMASKHYLSATDMLVSAVESLEGPLLQVEGL SDLRLELHSKKMNLHLVLIDELHRHLYIKSTSRVVQRNKEKGKISSLVKDASVPLIDV TNLPTPRKFLDTSHYSTAGSSSVREINLQDIKEDLELDPEENSTLFMGILIKGLAKLK KIPETVKAIIERLEQELKQIVKRSTTQVADSGYQRGENVTVENQPRLLLELLELLFDK FNAVAAAHSVVLGYLQDTVVTPLTQQEDIKLYDMADVWVKIQDVLQMLLTEYLDMKNT RTASEPSAQLSYASTGREFAAFFAKKKPQRPKNSLFKFESSSHAISMSAYLREQRREL YSRSGELQGGPDDNLIEGGGTKFVCKPGARNITVIFHPLLRFIQEIEHALGLGPAKQC PLREFLTVYIKNIFLNQVLAEINKEIEGVTKTSDPLKILANADTMKVLGVQRPLLQST IIVEKTVQDLLNLMHDLSAYSDQFLNMVCVKLQEYKDTCTAAYRGIVQSEEKLVISAS WAKDDDISRLLKSLPNWMNMAQPKQLRPKREEEEDFIRAAFGKESEVLIGNLGDKLIP PQDILRDVSDLKALANMHESLEWLASRTKSAFSNLSTSQMLSPAQDSHTNTDLPPVSE QIMQTLSELAKSFQDMADRCLLVLHLEVRVHCFHYLIPLAKEGNYAIVANVESMDYDP LVVKLNKDISAIEEAMSASLQQHKFQYIFEGLGHLISCILINGAQYFRRISESGIKKM CRNIFVLQQNLTNITMSREADLDFARQYYEMLYNTADELLNLVVDQGVKYTELEYIHA LTLLHRSQTGVGELTTQNTSCRGGSKRSSASRLPSSKPPRTRR

Further analysis of the NOV53a protein yielded the following properties shown in
Table 53B.

Table 53B. Protein Sequence Properties NOV53a	
PSort analysis:	0.5500 probability located in endoplasmic reticulum (membrane); 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	No Known Signal Sequence Predicted .

A search of the NOV53a protein against the Geneseq database, a proprietary
database that contains sequences published in patents and patent publication, yielded
5   several homologous proteins shown in Table 53C.

Table 53C. Geneseq Results for NOV53a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV53a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB93175	Human protein sequence SEQ ID NO:12114 - Homo sapiens, 974 aa. [EP1074617-A2, 07-FEB-2001]	1..947 1..947	947/947 (100%) 947/947 (100%)	0.0
AAW69801	Amino acid sequence of rsec8, a protein present in SA-17S complex - Rattus sp, 975 aa. [WO9828419-A2, 02-JUL-1998]	1..947 1..948	902/948 (95%) 925/948 (97%)	0.0
AAB95143	Human protein sequence SEQ ID NO:17163 - Homo sapiens, 572 aa. [EP1074617-A2, 07-FEB-2001]	403..947 1..545	545/545 (100%) 545/545 (100%)	0.0
AAB58175	Lung cancer associated polypeptide sequence SEQ ID 513 - Homo sapiens, 418 aa. [WO200055180-A2, 21-SEP-2000]	571..947 15..391	369/377 (97%) 369/377 (97%)	0.0
AAG00950	Human secreted protein, SEQ ID NO: 5031 - Homo sapiens, 100 aa. [EP1033401-A2, 06-SEP-2000]	451..544 7..100	76/94 (80%) 79/94 (83%)	3e-36

In a BLAST search of public sequence databases, the NOV53a protein was found to have homology to the proteins shown in the BLASTP data in Table 53D.

Table 53D. Public BLASTP Results for NOV53a				
Protein Accession Number	Protein/Organism/Length	NOV53a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96A65	CDNA FLJ14782 FIS, CLONE NT2RP4000524, HIGHLY SIMILAR TO MUS MUSCULUS SEC8 MRNA (SECRETORY PROTEIN SEC8) - Homo sapiens (Human), 974 aa.	1..947 1..947	947/947 (100%) 947/947 (100%)	0.0
Q9C0G4	KIAA1699 PROTEIN - Homo sapiens (Human), 966 aa (fragment).	9..947 1..939	939/939 (100%) 939/939 (100%)	0.0
O35382	SEC8 - Mus musculus (Mouse), 971 aa.	1..971 1..971	923/972 (94%) 946/972 (96%)	0.0

255

Q62824	RSEC8 - Rattus norvegicus (Rat), 975 aa (fragment).	1..947 1..948	902/948 (95%) 925/948 (97%)	0.0
Q9P102	REC8 - Homo sapiens (Human), 637 aa (fragment).	339..947 2..610	609/609 (100%) 609/609 (100%)	0.0

PFam analysis predicts that the NOV53a protein contains the domains shown in the Table 53E.

<table>
<tr><th colspan="4">Table 53E. Domain Analysis of NOV53a</th></tr>
<tr><th>Pfam Domain</th><th>NOV53a Match Region</th><th>Identities/
Similarities
for the Matched Region</th><th>Expect Value</th></tr>
<tr><td colspan="4">No Significant Matches Found</td></tr>
</table>

Example 54.

The NOV54 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 54A.

<table>
<tr><th colspan="4">Table 54A. NOV54 Sequence Analysis</th></tr>
<tr><td></td><td>SEQ ID NO: 161</td><td colspan="2">501 bp</td></tr>
<tr><td>NOV54a,
CG59545-01 DNA Sequence</td><td colspan="3">CAACACGAGGAACAATGTCTTCTTTACCCGTGCCATACAAACTGCCTGTGTCTTTGTC
TGTTGGTTCCTGCGTGATAATCAAAGGGACACTGATCGACTCTTCTATCAACGAACCA
CAGCTGCAGGTGGATTTCTACACTGAGATGAATGAGGACTCAGAAATTGCCTTCCATT
TGCGAGTGCACTTAGGCCGTCGTGTGGTCATGAACAGTCGTGAGTTTGGGATATGGAT
GTTGGAGGAGAATTTACACTATGTGCCCTTTGAGGATGGCAAACCATTTGACTTGCGC
ATCTACGTGTGTCTCAATGAGTATGAGGTAAAGGTAAATGGTGAATACATTTATGCCT
TTGTCCATCGAATCCCGCCATCATATGTGAAGATGATTCAAGTGTGGAGAGATGTCTC
CCTGGACTCAGTGCTTGTCAACAATGGACGGAGATGATCACACTCCTCATTGTTGAGG
AAACCCTCTTTCTACCTGACCATGGGATTCCTAGAGC</td></tr>
<tr><td></td><td>ORF Start: ATG at 15</td><td colspan="2">ORF Stop: TGA at 441</td></tr>
<tr><td></td><td>SEQ ID NO: 162</td><td>142 aa</td><td>MW at 16511.9kD</td></tr>
<tr><td>NOV54a,
CG59545-01 Protein Sequence</td><td colspan="3">MSSLPVPYKLPVSLSVGSCVIIKGTLIDSSINEPQLQVDFYTEMNEDSEIAFHLRVHL
GRRVVMNSREFGIWMLEENLHYVPFEDGKPFDLRIYVCLNEYEVKVNGEYIYAFVHRI
PPSYVKMIQVWRDVSLDSVLVNNGRR</td></tr>
</table>

Further analysis of the NOV54a protein yielded the following properties shown in Table 54B.

<table>
<tr><th colspan="2">Table 54B. Protein Sequence Properties NOV54a</th></tr>
<tr><td>PSort
analysis:</td><td>0.5500 probability located in endoplasmic reticulum (membrane); 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside</td></tr>
<tr><td>SignalP
analysis:</td><td>No Known Signal Sequence Predicted</td></tr>
</table>

A search of the NOV54a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 54C.

Table 54C. Geneseq Results for NOV54a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV54a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG66741	Human Charcot-Leyden crystal protein 5A (CLC5A) - Homo sapiens, 142 aa. [CN1302875-A, 11-JUL-2001]	1..142 1..142	139/142 (97%) 139/142 (97%)	2e-77
AAG66742	Human Charcot-Leyden crystal protein 5B (CLC5B) - Homo sapiens, 170 aa. [CN1302875-A, 11-JUL-2001]	6..142 34..170	136/137 (99%) 136/137 (99%)	3e-76
AAM79041	Human protein SEQ ID NO 1703 - Homo sapiens, 139 aa. [WO200157190-A2, 09-AUG-2001]	1..139 1..139	107/139 (76%) 116/139 (82%)	2e-56
AAY28350	Full Placental Protein 13 amino acid sequence - Homo sapiens, 139 aa. [WO9938970-A1, 05-AUG-1999]	1..139 1..139	107/139 (76%) 116/139 (82%)	2e-56
AAG78627	Human Charcot-Leyden crystal 4 CLC4 protein #2 - Homo sapiens, 167 aa. [CN1302876-A, 11-JUL-2001]	6..139 34..167	102/134 (76%) 111/134 (82%)	2e-53

In a BLAST search of public sequence databases, the NOV54a protein was found to have homology to the proteins shown in the BLASTP data in Table 54D.

Table 54D. Public BLASTP Results for NOV54a				
Protein Accession Number	Protein/Organism/Length	NOV54a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UHV8	PLACENTAL PROTEIN 13 (PLACENTA PROTEIN 13) - Homo sapiens (Human), 139 aa.	1..139 1..139	107/139 (76%) 116/139 (82%)	9e-56

257

Q9NR03	PLACENTAL PROTEIN 13-LIKE PROTEIN - Homo sapiens (Human), 139 aa.	1..139 1..139	86/139 (61%) 107/139 (76%)	9e-45
A46523	Charcot-Leyden crystal protein - human, 142 aa.	1..142 1..142	76/142 (53%) 96/142 (67%)	7e-36
Q05315	Eosinophil lysophospholipase (EC 3.1.1.5) (Charcot-Leyden crystal protein) (Lysolecithin acylhydrolase) (CLC) (Galactin-10) - Homo sapiens (Human), 141 aa.	2..142 1..141	75/141 (53%) 95/141 (67%)	3e-35
Q96KD6	PLACENTAL PROTEIN 13-LIKE - Homo sapiens (Human), 104 aa (fragment).	1..104 1..104	66/104 (63%) 79/104 (75%)	1e-31

PFam analysis predicts that the NOV54a protein contains the domains shown in the Table 54E.

Table 54E. Domain Analysis of NOV54a			
Pfam Domain	NOV54a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Gal-bind_lectin: domain 1 of 1	5..137	37/142 (26%) 106/142 (75%)	3.1e-28

Example 55.

The NOV55 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 55A.

Table 55A. NOV55 Sequence Analysis		
	SEQ ID NO: 163	2071 bp
NOV55a, CG59435-01 DNA Sequence	AAACTATTTGTAGGCGCAGTCATGCAGGAAAACCTCAGATTTGCTTCATCAGGAGATG ATATTAAAATATGGGATGCTTCATCTATGACATTGGTGGATAAATTCAACCCACACAC ATCACCACATGGAATCAGCTCAATATGTTGGAGCAGCAATAGTAACTTTTTAGTAACA GCATCTTCCAGTGGCGACAAAATAGTTGTCTCAAGTTGCAAATGTAAACCTGTTCCAC TTTTAGAGCTTGCTGAAGGGCAAAAGCAGACATGTGTCAATTTAAATTCTACATCTAT GTATTTGGTAAGCGGAGGCCTAAATAACACTGTTAATATTTGGGATTTAAAATCAAAA AGAGTTCATCGATCTCTTAAGGATCATAAAGATCAAGTAACTTGTGTAACATACAATT GGAATGATTGCTACATTGCTTCTGGATCTCTTAGTGGTGAAATTATTTTACACAGTGT AACCACTAATTTATCTAGTACTCCTTTTGGCCATGGTAGTAACCAGGTTCGGCACTTG AAGTACTCCTTGTTTAAGAAATCACTACTGGGCAGTGTTTCGGATAATGGAATAGTAA CTCTCTGGGATGTAAATAGTCAGAGTCCATACCATAACTTTGACAGTGTACACAAAGC TCCAGCGTCAGGCATCTGTTTTTCTCCTGTCAATGAATTGCTCTTTGTAACCATAGGC TTGGATAAAAGAATCATCCTCTATGACACTTCAAGTAAGAAGCTAGTGAAAACTTTAG TGGCTGACACTCCTCTAACTGCGGTAGATTTCATGCCTGATGGAGCCACTTTGGCTAT TGGATCTTCCCGGGGGAAAATATATCAATATGATTTAAGAATGTTGAAATCACCAGTT AAGACCATCAGTGCTCACAAGACATCTGTGCAGTGTATAGCATTTCAGTACTCCACTG TTCTTACTAAGTCAAGTTTAAATAAAGGCTGTTCAAATAAGCCCACAACAGTGAACAA ACGAAGTGTTAATGTGAATGCTGCTAGTGGAGGAGTTCAGAATTCCGGAATTGTCAGA GAAGCACCTGCCACGTCCATTGCCACAGTTCTACCACAACCTATGACATCAGCTATGG

258

	GGAAAGGAACAGTTGCTGTTCAAGAAAAAGCAGGTTTGCCTCGAAGCATAAACACAGA CACTTTATCTAAGGAAACAGACAGTGGAAAAAATCAGGATTTCTCCAGCTTTGATGAT ACTGGGAAAAGTAGTTTAGGTGACATGTTCTCACCTATCAGAGATGATGCTGTAGTTA ACAAGGGAAGTGATGAGTCCATAGGCAAAGGAGATGGCTTTGACTTTCTACCGCAGTT GAACTCAGTGTTTCCTCCAAGAAAAAATCCAGTAACTTCAAGTACTTCAGTATTGCAT TCTAGTCCTCTTAATGTTTTTATGGGATCTCCAGGGAAAGAGGAAAATGAAACCGTG ATCTAACAGCTGAGTCTAAGAAAATATATATGGGAAAACAGGAATCTAAAGACTCCTT CAAACAGTTAGCAAAGTTGGTCACATCTGGTGCTGAAAGTGGAAATCTAAATACCTCT CCATCATCTAACCAAACAAGAAATTCTGAGAAATTTGAAAAGCCAGAGAATGAAATTG AAGCCCAGTTGATATGTGAACCCCCAATCAATGGATCCTCAACTCCAAATCCAAAGAT AGCATCTTCTGTCACTGCTGGAGTTGCCAGTTCACTCTCAGAAAAAATAGCCGACAGC ATTGGAAATAACCGGCAAAATGCACCATTGACTTCCATTCAAATTCGTTTTATTCAGA ACATGATACAGGAAACGTTGGATGACTTTAGAGAAGCATGCCATAGGGACATTGTGAA TTTGCAAGTGGAGATGATTAAACAGTTTCATATGCAACTGAATGAAATGCATTCTTTG CTGGAAAGATACTCAGTGAATGAAGGTTTAGTGGCTGAAATTGAAAGACTACGAGAAG AAAACAAAAGATTACGGGCCCACTTT**TGA**AATTTCAGTGAATACCTTAATGTTCTGTA ATTTGGGAAGTTTCTGGCAACACAGAACTACATAGAATCAT
	ORF Start: ATG at 22	ORF Stop: TGA at 1999
	SEQ ID NO: 164	659 aa	MW at 71851.2kD

NOV55a, CG59435-01 Protein Sequence	MQENLRFASSGDDIKIWDASSMTLVDKFNPHTSPHGISSICWSSNSNFLVTASSSGDK IVVSSCKCKPVPLLELAEGQKQTCVNLNSTSMYLVSGGLNNTVNIWDLKSKRVHRSLK DHKDQVTCVTYNWNDCYIASGSLSGEIILHSVTTNLSSTPFGHGSNQVRHLKYSLFKK SLLGSVSDNGIVTLWDVNSQSPYHNFDSVHKAPASGICFSPVNELLFVTIGLDKRIIL YDTSSKKLVKTLVADTPLTAVDFMPDGATLAIGSSRGKIYQYDLRMLKSPVKTISAHK TSVQCIAFQYSTVLTKSSLNKGCSNKPTTVNKRSVNVNAASGGVQNSGIVREAPATSI ATVLPQPMTSAMGKGTVAVQEKAGLPRSINTDTLSKETDSGKNQDFSSFDDTGKSSLG DMFSPIRDDAVVNKGSDESIGKGDGFDFLPQLNSVFPPRKNPVTSSTSVLHSSPLNVF MGSPGKEENENRDLTAESKKIYMGKQESKDSFKQLAKLVTSGAESGNLNTSPSSNQTR NSEKFEKPENEIEAQLICEPPINGSSTPNPKIASSVTAGVASSLSEKIADSIGNNRQN APLTSIQIRFIQNMIQETLDDFREACHRDIVNLQVEMIKQFHMQLNEMHSLLERYSVN EGLVAEIERLREENKRLRAHF
	SEQ ID NO: 165	2009 bp

NOV55b, CG59435-02 DNA Sequence	AAACTATTTGTAGGCGCAGTCATGCAGGAAAACCTCAGATTTGCTTCATCAGGAGATG ATATTAAAATATGGGATGCTTCATCTATGACATTGGTGGATAAATTCAACCCACACAC ATCACCACATGGAATCAGCTCAATATGTTGGAGCAGCAATAATAACTTTTTAGTAACA GCATCTTCCAGTGGCGACAAAATAGTTGTCTCAAGTTGCAAATGTAAACCTGTTCCAC TTTTAGAGCTTGCTGAAGGGCAAAAGCAGACATGTGTCAATTTAAATTCTACATCTAT GTATTTGGTAAGCGGAGGCCTAAATAACACTGTTAATATTTGGGATTTAAAATCAAAA AGAGTTCATCGATCTCTTAAGGATCATAAAGATCAAGTAACTTGTGTAACATACAATT GGAATGATTGCTACATTGCTTCTGGATCTCTTAGTGGTGAAATTATTTTACACAGTGT AACCACTAATTTATCTAGTACTCCTTTTGGCCATGGTAGTAACCAGTCTGTTCGGCAC TTGAAGTACTCCTTGTTTAAGAAATCACTACTGGGCAGTGTTTCGGATAATGGAATAG TAACTCTCTGGGATGTAAATAGTCAGAGTCCATACCATAACTTTGACAGTGTACACAA AGCTCCAGCGTCAGGCATCTGTTTTTCTCCTGTCAATGAATTGCTCTTTGTAACCATA GGCTTGGATAAAAGAATCATCCTCTATGACACTTCAAGTAAGAAGCTAGTGAAAACTT TAGTGGCTGACACTCCTCTAACTGCGGTAGATTTCATGCCTGATGGAGCCACTTTGGC TATTGGATCTTCCCGGGGGAAAATATATCAATATGATTTAAGAATGTTGAAATCACCA GTTAAGACCATCAGTGCTCACAAGACATCTGTGCAGTGTATAGCATTTCAGTACTCCA CTGTTCTTACTAAGTCAAGTTTAAATAAAGGCTGTTCAAATAAGCCCACAACAGTGAA CAAACGAAGTGTTAATGTGAATGCTGCTAGTGGAGGAGTTCAGAATTCCGGAATTGTC AGAGAAGCACCTGCCACGTCCATTGCCACAGTTCTACCACAACCTATGACATCAGCTA TGGGGAAAGGAACAGTTGCTGTTCAAGAAAAAGCAGGTTTGCCTCGAAGCATAAACAC AGACACTTTATCTAAGGAAACAGACAGTGGAAAAAATCAGGATTTCTCCAGCTTTGAT GATACTGGGAAAAGTAGTTTAGGTGACATGTTCTCACCTATCAGAGATGATGCTGTAG TTAACAAGGGAAGTGATGAGTCCATAGGCAAAGGAGATGGCTTTGACTTTCTACCGCA GTTGAACTCAGTGTTTCCTCCAAGAAAAAATCCAGTAACTTCAAGTACTTCAGTATTG CATTCTAGTCCTCTTAATGTTTTTATGGGATCTCCAGGGAAAGAGGAAAATGAAACC GTGATCTAACAGCTGAGTCTAAGAAAATATATATGGGAAAACAGGAATCTAAAGACTC CTTCAAACAGTTAGCAAAGTTGGTCACATCTGGTGCTGAAAGTGGAAATCTAAATACC TCTCCATCATCTAACCAAACAAGAAATTCTGAGAAATTTGAAAAGCCAGAGAATGAAA TTGAAGCCCAGTTGATATGTGAACCCCCAATCAATGGATCCTCAACTCCAAATCCAAA GATAGCATCTTCTGTCACTGCTGGAGTTGCCAGTTCACTCTCAGAAAAAATAGCCGAC AGCATTGGAAATAACCGGCAAAATGCACCATTGACTTCCATTCAAATTCGTTTTATTC AGAACATGATACAGGAAACGTTGGATGACTTTAGAGAAGCATGCCATAGGGACATTGT GAATTTGCAAGTGGAGATGATTAAACAGTTTCATATGCAACTGAATGAAATGCATTCT TTGCTGGAAAGATACTCAGTGAATGAAGGTTTAGTGGCTGAAATTGAAAGACTACGAG AAGAAAACAAAAGATTACGGGCCCACTTT**TGAAATTT**
	ORF Start: ATG at 22	ORF Stop: TGA at 2002
	SEQ ID NO: 166	660 aa	MW at 71965.3kD
NOV55b,	MQENLRFASSGDDIKIWDASSMTLVDKFNPHTSPHGISSICWSSNNNFLVTASSSGDK

259

CG59435-02 Protein Sequence	IVVSSCKCKPVPLLELAEGQKQTCVNLNSTSMYLVSGGLNNTVNIWDLKSKRVHRSLK
	DHKDQVTCVTYNWNDCYIASGSLSGEIILHSVTTNLSSTPFGHGSNQSVRHLKYSLFK
	KSLLGSVSDNGIVTLWDVNSQSPYHNFDSVHKAPASGICFSPVNELLFVTIGLDKRII
	LYDTSSKKLVKTLVADTPLTAVDFMPDGATLAIGSSRGKIYQYDLRMLKSPVKTISAH
	KTSVQCIAFQYSTVLTKSSLNKGCSNKPTTVNKRSVNVNAASGGVQNSGIVREAPATS
	IATVLPQPMTSAMGKGTVAVQEKAGLPRSINTDTLSKETDSGKNQDFSSFDDTGKSSL
	GDMFSPIRDDAVVNKGSDESIGKGDGFDFLPQLNSVFPPRKNPVTSSTSVLHSSPLNV
	FMGSPGKEENENRDLTAESKKIYMGKQESKDSFKQLAKLVTSGAESGNLNTSPSSNQT
	RNSEKFEKPENEIEAQLICEPPINGSSTPNPKIASSVTAGVASSLSEKIADSIGNNRQ
	NAPLTSIQIRFIQNMIQETLDDFREACHRDIVNLQVEMIKQFHMQLNEMHSLLERYSV
	NEGLVAEIERLREENKRLRAHF

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 55B.

Table 55B. Comparison of NOV55a against NOV55b.		
Protein Sequence	NOV55a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV55b	1..659 1..660	658/660 (99%) 659/660 (99%)

Further analysis of the NOV55a protein yielded the following properties shown in Table 55C.

Table 55C. Protein Sequence Properties NOV55a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV55a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 55D.

Table 55D. Geneseq Results for NOV55a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV55a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG74568	Human colon cancer antigen protein SEQ ID NO:5332 - Homo sapiens, 404 aa. [WO200122920-A2, 05-APR-2001]	256..659 1..404	399/404 (98%) 399/404 (98%)	0.0

260

AAE10677	Human NEDD1-related protein - Homo sapiens, 208 aa. [WO200172955-A2, 04-OCT-2001]	453..611 2..159	145/159 (91%) 149/159 (93%)	4e-75
AAM70774	Human bone marrow expressed probe encoded protein SEQ ID NO: 31080 - Homo sapiens, 67 aa. [WO200157276-A2, 09-AUG-2001]	240..306 1..67	67/67 (100%) 67/67 (100%)	9e-31
AAM06190	Peptide #4872 encoded by probe for measuring breast gene expression - Homo sapiens, 67 aa. [WO200157270-A2, 09-AUG-2001]	240..306 1..67	67/67 (100%) 67/67 (100%)	9e-31
ABB23122	Protein #5121 encoded by probe for measuring heart cell gene expression - Homo sapiens, 65 aa. [WO200157274-A2, 09-AUG-2001]	307..371 1..65	65/65 (100%) 65/65 (100%)	3e-29

In a BLAST search of public sequence databases, the NOV55a protein was found to have homology to the proteins shown in the BLASTP data in Table 55E.

Table 55E. Public BLASTP Results for NOV55a				
Protein Accession Number	Protein/Organism/Length	NOV55a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
I60167	regulatory protein Nedd1 - mouse, 660 aa.	1..659 1..660	564/660 (85%) 607/660 (91%)	0.0
P33215	NEDD1 protein - Mus musculus (Mouse), 675 aa (fragment).	1..659 16..675	564/660 (85%) 607/660 (91%)	0.0
Q9CWK2	NEURAL PRECURSOR CELL EXPRESSED, DEVELOPMENTALLY DOWN-REGULATED GENE 1 - Mus musculus (Mouse), 660 aa.	1..659 1..660	563/660 (85%) 606/660 (91%)	0.0
Q9FI89	SIMILARITY TO REGULATORY PROTEIN NEDD1 - Arabidopsis thaliana (Mouse-ear cress), 787 aa.	8..533 15..532	145/550 (26%) 246/550 (44%)	4e-40
BAB75165	WD-40 REPEAT PROTEIN - Anabaena sp. (strain PCC 7120), 1526 aa.	2..298 916..1208	92/307 (29%) 147/307 (46%)	2e-18

PFam analysis predicts that the NOV55a protein contains the domains shown in the Table 55F.

5   261

Table 55F. Domain Analysis of NOV55a			
Pfam Domain	**NOV55a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
WD40: domain 1 of 7	28..61	6/37 (16%) 27/37 (73%)	57
WD40: domain 2 of 7	70..105	10/37 (27%) 27/37 (73%)	0.062
WD40: domain 3 of 7	111..147	9/37 (24%) 28/37 (76%)	20
WD40: domain 4 of 7	153..190	10/38 (26%) 29/38 (76%)	3.4
WD40: domain 5 of 7	197..234	7/38 (18%) 25/38 (66%)	19
WD40: domain 6 of 7	240..275	14/37 (38%) 28/37 (76%)	3.1
WD40: domain 7 of 7	282..316	8/37 (22%) 26/37 (70%)	1.3e+03

Example 56.

The NOV56 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 56A.

Table 56A. NOV56 Sequence Analysis		
	SEQ ID NO: 167	1771 bp
NOV56a, CG59439-01 DNA Sequence	GACTGTTTCACCATGCAGTGGCTAATGAGGTTCCGGACCCTCTGGGGCATCCACAAAT CCTTCCACAACATCCACCCTGCCCCTTCACAGCTGCGCTGCCGGTCTTTATCAGAATT TGGAGCCCCAAGATGGAATGACTATGAAGTACCGGAGGAATTTAACTTTGCAAGTTAT GTACTGGACTACTGGGCTCAAAAGGAGAAGGAGGGCAAGAGAGGTCCAAATCCAGCTT TTTGGTGGGTGAATGGCCAAGGGGATGAAGTAAAGTGGAGCTTCAGAGAGATGGGAGA CCTAACCCGCCGTGTAGCCAACGTCTTCACACAGACCTGTGGCCTACAACAGGGAGAC CATCTGGCCTTGATGCTGCCTCGAGTTCCTGAGTGGTGGCTGGTGGCTGTGGGCTGCA TGCGAACAGGGATCATCTTCATTCCTGCGACCATCCTGTTGAAGGCCAAAGACATTCT CTATCGACTACAGTTGTCTAAAGCCAAGGGCATTGTGACCATAGATGCCCTTGCCTCA GAGGTGGACTCCATAGCTTCTCAGTGCCCCTCTCTGAAAACCAAGCTCCTGGTGTCTG ATCACAGCCGTGAAGGGTGGCTGGACTTCCGATCGCTGGTTAAATCAGCATCCCCAGA ACACACCTGTGTTAAGTCAAAGACCTTGGACCCAATGGTCATCTTCTTCACCAGTGGG ACCACAGGCTTCCCCAAGATGGCAAAACACTCCCATGGGTTGGCCTTACAACCCTCCT TCCCAGGAAGTAGGAAATTACGGAGCCTGAAGACATCTGATGTCTCCTGGTGCCTGTC GGACTCAGGATGGATTGTGGCTACCATTTGGACCCTGGTAGAACCATGGACAGCGGGT TGTACAGTCTTTATCCACCATCTGCCACAGTTTGACACCAAGGTCATCATACAGACAT TGTTGAAATACCCCATTAACCACTTTTGGGGGGTATCATCTATATATCGAATGATTCT GCAGCAGGATTTCACCAGCATCAGGTTCCCTGCCCTGGAGCACTGCTATACTGGCGGG GAGGTCGTGTTGCCCAAGGATCAGGAGGAGTGGAAAAGACGGACGGGCCTTCTGCTCT ACGAGAACTATGGGCAGTCGGAAACGGGACTAATTTGTGCCACCTACTGGGGAATGAA GATCAAGCCGGGTTTCATGGGGAAGGCCACTCCACCCTACGACGTCCAGGTCATTGAT GACAAGGGCAGCATCCTGCCACCTAACACAGAAGGAAACATTGGCATCAGAATCAAAC CTGTCAGGCCTGTGAGCCTCTTCATGTGCTATGAGGGTGACCCAGAGAAGACAGCTAA AGTGGAATGTGGGGACTTCTACAACACTGGGGACAGAGGTAAGATGGATGAAGAGGGC TACATTTGTTTCCTGGGGAGGAGTGATGACATCATTAATGCCTCTGGGTATCGCATCG GGCCTGCAGAGGTTGAAAGCGCTTTGGTGGAGCACCCAGCGGTGGCGGAGTCAGCCGT GGTGGGCAGCCCAGACCCGATTCGAGGGGAGGTGGTGAAGGCCTTTATTGTCCTGACC

262

	CCACAGTTCCTGTCCCATGACAAGGATCAGCTGACCAAGGAACTGCAGCAGCATGTCA AGTCAGTGACAGCCCCATACAAGTACCCAAGGAAGGTGGAGTTTGTCTCAGAGCTGCC AAAAACCATCACTGGCAAGATTGAACGGAAGGAACTTCGGAAAAAGGAGACTGGTCAG ATG**TAA**TCGGCAGTGAACTCAGAACGCACTG		
	ORF Start: ATG at 13	ORF Stop: TAA at 1744	
	SEQ ID NO: 168	577 aa	MW at 65272.6kD
NOV56a, CG59439-01 Protein Sequence	MQWLMRFRTLWGIHKSFHNIHPAPSQLRCRSLSEFGAPRWNDYEVPEEFNFASYVLDY WAQKEKEGKRGPNPAFWWVNGQGDEVKWSFREMGDLTRRVANVFTQTCGLQQGDHLAL MLPRVPEWWLVAVGCMRTGIIFIPATILLKAKDILYRLQLSKAKGIVTIDALASEVDS IASQCPSLKTKLLVSDHSREGWLDFRSLVKSASPEHTCVKSKTLDPMVIFFTSGTTGF PKMAKHSHGLALQPSFPGSRKLRSLKTSDVSWCLSDSGWIVATIWTLVEPWTAGCTVF IHHLPQFDTKVIIQTLLKYPINHFWGVSSIYRMILQQDFTSIRFPALEHCYTGGEVVL PKDQEEWKRRTGLLLYENYGQSETGLICATYWGMKIKPGFMGKATPPYDVQVIDDKGS ILPPNTEGNIGIRIKPVRPVSLFMCYEGDPEKTAKVECGDFYNTGDRGKMDEEGYICF LGRSDDIINASGYRIGPAEVESALVEHPAVAESAVVGSPDPIRGEVVKAFIVLTPQFL SHDKDQLTKELQQHVKSVTAPYKYPRKVEFVSELPKTITGKIERKELRKKETGQM		
	SEQ ID NO: 169	1659 bp	
NOV56b, CG59439-02 DNA Sequence	GTTTCACC**ATG**CAGTGGCTAATGAGGTTCCGGACCCTCTGGGGCATCCACAAATCCTT CCACAACATCCACCCTGCCCCTTCACAGCTGCGCTGCCGGTCTTTATCAGAATTTGGA GCCCCAAGATGGAATGACTATGAAGTACCGGAGGAATTTAACTTTGCAAGTTATGTAC TGGACTACTGGGCTCAAAAGGAGAAGGAGGGCAAGAGAGGTCCAAATCCAGCTTTTTG GTGGGTGAATGGCCAAGGGGATGAAGTAAAGTGGAGCTTCAGAGAGATGGGAGACCTA ACCCGCCGTGTAGCCAACGTCTTCACACAGACCTGTGGCCTACAACAGGGAGACCATC TGGCCTTGATGCTGCCTCGAGTTCCTGAGTGGTGGCTGGTGGCTGTGGGCTGCATGCG AACAGGGATCATCTTCATTCCTGCGACCATCCTGTTGAAGGCCCAAAGACATTCTCTAT CGACTACAGTTGTCTAAAGCCAAGGGCATTGTGACCATAGATGCCCTTGCCTCAGAGG TGGACTCCATAGCTTCTCAGTGCCCCTCTCTGAAAACCAAGCTCCTGGTGTCTGATCA CAGCCGTGAAGGGTGGCTGGACTTCCGATCGCTGGTTAAATCAGCATCCCCAGAACAC ACCTGTGTTAAGTCAAAGACCTTGGACCCAATGGTCATCTTCTTCACCAGTGGGACCA CAGGCTTCCCCAAGATGGCAAAACACTCCCATGGGTTGGCCTTACAACCCTCCTTCCC AGGAAGTAGGAAATTACGGAGCCTGAAGACATCTGATGTCTCCTGGTGCCTGTCGGAC TCAGGATGGATTGTGGCTACCATTTGGACCCTGGTAGAACCATGGACAGCGGGTTGTA CAGTCTTTATCCACCATCTGCCACAGTTTGACACCAAGGTCATCATACAGACATTGTT GAAATACCCCATTAACCACTTTTGGGGGGTATCATCTATATATCGAATGATTCTGCAG CAGGATTTCACCAGCATCAGGTTCCCTGCCCTGGAGCACTGCTATACTGGCGGGGAGG TCGTGTTGCCCAAGGATCAGGAGGAGTGGAAAAGACGGACGGGCCTTCTGCTCTACGA GAACTATGGGCAGTCGGAAACGGGACTAATTTGTGCCACCTACTGGGGAATGAAGATC AAGCCGGGTTTCATGGGGAAGGCCACTCCACCCTACGACGTCCAGGGTGACCCAGAGA AGACAGCTAAAGTGGAATGTGGGGACTTCTACAACACTGGGGACAGAGGAAAGATGGA TGAAGAGGGCTACATTTGTTTCCTGGGGAGGAGTGATGACATCATTAATGCCTCTGGG TATCGCATCGGGCCTGCAGAGGTTGAAAGTGCTTTGGTGGAGCACCCAGCGGTGGCGG AGTCAGCCGTGGTGGGCAGCCCAGACCCGATTCGAGGGGAGGTGGTGAAGGCCTTTAT TGTCCTGACCCCACAGTTCCTGTCCCATGACAAGGATCAGCTGACCAAGGAACTGCAG CAGCATGTCAAGTCAGTGACAGCCCCATACAAGTACCCAAGGAAAGTGGAGTTTGTCT CAGAGCTGCCAAAAACCATCACTGGCAAGATTGAACGGAAGGAACTTCGGAAAAAGGA GACTGGTCAGATG**TAA**TCGGCAGTGAACTCAGAAC		
	ORF Start: ATG at 9	ORF Stop: TAA at 1638	
	SEQ ID NO: 170	543 aa	MW at 61518.2kD
NOV56b, CG59439-02 Protein Sequence	MQWLMRFRTLWGIHKSFHNIHPAPSQLRCRSLSEFGAPRWNDYEVPEEFNFASYVLDY WAQKEKEGKRGPNPAFWWVNGQGDEVKWSFREMGDLTRRVANVFTQTCGLQQGDHLAL MLPRVPEWWLVAVGCMRTGIIFIPATILLKAKDILYRLQLSKAKGIVTIDALASEVDS IASQCPSLKTKLLVSDHSREGWLDFRSLVKSASPEHTCVKSKTLDPMVIFFTSGTTGF PKMAKHSHGLALQPSFPGSRKLRSLKTSDVSWCLSDSGWIVATIWTLVEPWTAGCTVF IHHLPQFDTKVIIQTLLKYPINHFWGVSSIYRMILQQDFTSIRFPALEHCYTGGEVVL PKDQEEWKRRTGLLLYENYGQSETGLICATYWGMKIKPGFMGKATPPYDVQGDPEKTA KVECGDFYNTGDRGKMDEEGYICFLGRSDDIINASGYRIGPAEVESALVEHPAVAESA VVGSPDPIRGEVVKAFIVLTPQFLSHDKDQLTKELQQHVKSVTAPYKYPRKVEFVSEL PKTITGKIERKELRKKETGQM		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 56B.

Table 56B. Comparison of NOV56a against NOV56b.		
Protein Sequence	NOV56a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV56b	1..577 1..543	543/577 (94%) 543/577 (94%)

Further analysis of the NOV56a protein yielded the following properties shown in Table 56C.

Table 56C. Protein Sequence Properties NOV56a	
PSort analysis:	0.6400 probability located in microbody (peroxisome); 0.4712 probability located in mitochondrial matrix space; 0.1737 probability located in mitochondrial inner membrane; 0.1737 probability located in mitochondrial intermembrane space
SignalP analysis:	Likely cleavage site between residues 23 and 24

A search of the NOV56a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 56D.

Table 56D. Geneseq Results for NOV56a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV56a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB43245	Human ORFX ORF3009 polypeptide sequence SEQ ID NO:6018 - Homo sapiens, 537 aa. [WO200058473-A2, 05-OCT-2000]	46..573 1..527	309/529 (58%) 402/529 (75%)	0.0
AAM80008	Human protein SEQ ID NO 3654 - Homo sapiens, 302 aa. [WO200157190-A2, 09-AUG-2001]	331..577 24..302	247/279 (88%) 247/279 (88%)	e-140
AAM80007	Human protein SEQ ID NO 3653 - Homo sapiens, 302 aa. [WO200157190-A2, 09-AUG-2001]	331..577 24..302	247/279 (88%) 247/279 (88%)	e-140

AAM41894	Human polypeptide SEQ ID NO 6825 - Homo sapiens, 390 aa. [WO200153312-A1, 26-JUL-2001]	257..573 7..323	193/317 (60%) 246/317 (76%)	e-116
AAM79024	Human protein SEQ ID NO 1686 - Homo sapiens, 196 aa. [WO200157190-A2, 09-AUG-2001]	382..577 1..196	196/196 (100%) 196/196 (100%)	e-112

In a BLAST search of public sequence databases, the NOV56a protein was found to have homology to the proteins shown in the BLASTP data in Table 56E.

Table 56E. Public BLASTP Results for NOV56a				
Protein Accession Number	Protein/Organism/Length	NOV56a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96A20	MIDDLE-CHAIN ACYL-COA SYNTHETASE1 (MEDIUM-CHAIN ACYL-COA SYNTHETASE) - Homo sapiens (Human), 577 aa.	1..577 1..577	576/577 (99%) 576/577 (99%)	0.0
Q9TVB5	XENOBIOTIC/MEDIUM-CHAIN FATTY ACID:COA LIGASE FORM XL-III PRECURSOR - Bos taurus (Bovine), 577 aa.	1..576 1..576	439/576 (76%) 486/576 (84%)	0.0
Q9BEA2	LIPOATE-ACTIVATING ENZYME PRECURSOR - Bos taurus (Bovine), 577 aa.	1..576 1..576	438/576 (76%) 485/576 (84%)	0.0
Q91VA0	MEDIUM-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.2) (HYPOTHETICAL 64.8 KDA PROTEIN) - Mus musculus (Mouse), 573 aa.	1..577 1..573	406/577 (70%) 472/577 (81%)	0.0
O70490	KIDNEY-SPECIFIC PROTEIN - Rattus norvegicus (Rat), 572 aa.	1..573 1..567	315/580 (54%) 417/580 (71%)	0.0

PFam analysis predicts that the NOV56a protein contains the domains shown in the

5   Table 56F.

265

Table 56F. Domain Analysis of NOV56a

Pfam Domain	NOV56a Match Region	Identities/ Similarities for the Matched Region	Expect Value
AMP-binding: domain 1 of 1	87..499	106/425 (25%) 299/425 (70%)	2.5e-96

Example 57.

The NOV57 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 57A.

Table 57A. NOV57 Sequence Analysis

	SEQ ID NO: 171	2501 bp	
NOV57a, CG59354-01 DNA Sequence	ACACCATGACCACCCTTGATGATAAGTTGCTGGGGGAGAAACTGCAGTACTACTATAG CAGCAGTGAGGATGAGGACAGTGACCACGAGGACAAGGACCGAGGCAGATGTGCCCCA GCCAGCAGTTCTGTGCCTGCAGAGGCTGAGCTGGCAGGCGAAGGCATCTCAGTTAACA CAATGACTCTGAAGGAGTTTGCCATAATGAATGAGGACCAAGATGATGAAGAGTTTCT GCAGCAGTACCGGAAGCAGCGAATGGAAGAGATGCGGCAGCAGCTTCACAAGGGGCCC CAATTCAAGCAGGTTTTTGAGATCTCCAGTGGAGAAGGGTTTTTAGACATGATTGATA AAGAACAGAAAAGCATTGTCATCATGGTTCATATTTATGAGGATGGCATTCCAGGGAC CGAAGCCATGAATGGTTGCATGATCTGCCTTGCCGCAGAGTACCCAGCTGTCAAGTTC TGCAAGGTGAAGAGCTCAGTTATTGGCGCCAGCAGTCAGTTCACCAGGAATGCCCTTC CTGCCCTGCTGATCTATAAGGGGGGTGAATTGATCGGCAATTTTGTTCGTGTTACTGA CCAGCTGGGGGATGATTTCTTTGCTGTGGACCTTGAAGCTTTTCTCCAGGAATTTGGA TTACTCCCAGAAAAGGAAGTCTTGGTGCTGACATCTGTGCGTAACTCTGCCACGTGTC ACAGTGAGGATAGCGACCTGGAAATAGATTGAACTGATAGTCTAGTTGCATAGATTTC TCATTGTTTGGGTTGGAATACACGTCATTGTTTATTTTTGTTCCTTTGTCTTCTGGCT TTTCAGCTGTTCTTTGTAGTCCCTTTTATTATGCATAAAATAAAGAAATTCTTAGATT AAATCAGAATGCTGAATAACCTTGTAGCTAGCAATAAGGTGACTTACAATTGTATAAA CAGGAAGCCAGGCTTTTGAACTGTTTACTTAAGATTCTGTGGTGTGACATCTCTGTTA TTGTTTCCAGTCAATATTTACAAAGCATCCTAAAGACAGGGTCTTGGAAATTGTCTTC AGATGATCTTAGAGGTCTCTGCCAAGTCTGAGAGTATAATTCTGTAGGTATTGTGTTA TTTGCAACGTAAATAGTGCATTTTCTTAATCAAATGATTGTAAATTATATTTACTTGT AATCAGTTCCATAGCTTTAGACGGTGGTTAGATTTTTTTTTTCCCCACCAGGGTCTTG TTTAAAGGGGTGAGCCACCGCACCCAGTCCTGAGGGGTGGCCTCTGCTGCTGGATTTC ATGTCTTCCTCCAGCATGACTAAGTCTGGAACAGCAGGAAGGGTTGATGCTTACTGAC CTGGTGATGTTAGAAGACAAGTAGTTTATGGATTTAAACATTAGAGCTGGAGTGGGGC TGGAAATCTTTGTAAAGGAAGTTCTTTCAGTAAGATGCCCCTGCTTGTCTTTGTCTCT TTTTTGTTTAACAAGGTAACTTTTTGTTTAACAAGGTAACTTTTTGTTTAACCTAGAT TTTTTTTAAAACTTTTTTTTTTTTTTCATATTGGAAAAGTAATTCATATTCAGTAGAGG AAAACTGACCAAAACAGAAGCAAAAATAAGAAAATTAAAATAATCTCTAATCCTACTA CCTAGAATAAAACACTATTAATATTTTGGTCTGTTTCCTGCCAAGGTGTTTTCTGTGT ATACATGGATATTTTGTTTGTTTTTAAACAAAACGATGGGATCATTCTGAACATACTG TTCTATAGTATGGTCAGCTAATAATATATCAGACCTTTTTTTTATATTATTAAATATT CTACAACTTTTTAAAAATGTCTATTAATATTCCATCGTATAGATGTGATATAATTTGC TTGATGGTTGTCTCTTAAAAAGAAAGATAGCAAATACTTTTTTTAAATTACAAAAGTG ATAGATGTTCATTGTAGAAAATGTAATAAACACTGTTAAGACTTAAAAGCCATATAAT TCCACCAACCAAAATTAATCCCTTTTGTCATATTTCTAGTCATTTTTATAGCCTTTTT TTTCTATGTATTTATAATAATTATCATTTGCGTTTTTTTCCTTTTTTTAACTTTAAAA ATGTATATTCTAGGGTCAGGGGAAATGTAATCTGGAATTAAATATTAGCCTTAAAATT CACAATTTTGATTTTCCTGGCTTTTCAGGAATTGACTAACTGTAAAAGAGTCTTGAAA GTATTTAGTCAACAAACAGAGTGCATTTTTTTTTTTTTTTGACTAAGAAAGCTCGTTGTA GTAGAAAGGGTGGAATGTATTGAAAATTATTAGAAGCAGGGAAGTATTGTTAGTCTAG CTTATTTCCTTTCAGTCTTTTTTCAATATTTTTATAAACATTGAGTACTTACTGAATT TAGTTCTGTGCTCTTCCTTATTTAGTGTTGTATCATAAATACTTTGATGTTTCAAACA TTCTAAATAAATAATTTTCAGTGGCTTCATAATAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAA	
	ORF Start: ATG at 6	ORF Stop: TGA at 726	
	SEQ ID NO: 172	240 aa	MW at 26866.9kD
NOV57a,	MTTLDDKLLGEKLQYYYSSSEDEDSDHEDKDRGRCAPASSSVPAEAELAGEGISVNTM TLKEFAIMNEDQDDEEFLQQYRKQRMEEMRQQLHKGPQFKQVFEISSGEGFLDMIDKE		

266

CG59354-01 Protein Sequence	QKSIVIMVHIYEDGIPGTEAMNGCMICLAAEYPAVKFCKVKSSVIGASSQFTRNALPA LLIYKGGELIGNFVRVTDQLGDDFFAVDLEAFLQEFGLLPEKEVLVLTSVRNSATCHS EDSDLEID
	SEQ ID NO: 173 · 893 bp
NOV57b, CG59354-02 DNA Sequence	CACCATGACCACCCTGTATGATAAGTTGCTGGGGGAGAAACTGCAGTACTACTATAGC AGCAGTGAAGATGAGGACAGTGACCACGAGGACAAGGACCGAGGCATCTCAGTTAACA CAGGCCCAAAAGGTGTGATCAATGACTGGCGCCGCTTCAAGCAGTTGGAGACAGAGCA GAGGGAGGAGCAGTGCCGGGAGATGGAAAGGCTGATCAAGAAGCTGTCAATGACTTGC AGGTCCCATCTGGATGAAGAGGAGGAGCAACAGAAACAGAAAGACCTCCAGGAGAAGA TCAGTGGGAAGATGACTCTGAAGGAGTTTGCCATAATGAATGAGGACCAAGATGATGA AGAGTTTCTGCAGCAGTACCGGAAGCAGCGAATGGAAGAGATGCGGCAGCAGCTTCAC AAGGGGCCCCAATTCAAGCAGGTTTTTGAGATCTCCAGTGGAGAAGGGTTTTTAGACA TGATTGATAAAGAACAGAAAAGCATTGTCATCATGGTTCATATTTATGAGGATGGCAT CAGGGACCGAAGCCATGAATGGTTGCATGATCCGCCTTGCAAGGGGGGTGAATTGATC GGCAATTTTGTTCGTGTTACTGACCAGCTGGGGGATGATTTCTTTGCTGTGGACCTTG AAGCTTTTCTCCAGGAATTTGGATTACTCCCAGAAAAGGAAGTCTTGGTGCTGACATC TGTGCGTAACTCTGCCACGTGTCACAGTGAGGATAGCGACCTGGAAATAGAT**TGA**ACT GATAGTCTAGTTGCATATAGATTTCTCATTGTTTGGGTTGGAATACACCATTGTTTAT TTTTGTTCCTTTGTCTTCTGGCTTTTCAGCTGTTCTTTGTAGTCCCTTTTATTATGCA TAAAATAAAGAAATTCTTAGATT
	ORF Start: ATG at 5 · **ORF Stop: TGA at 749**
	SEQ ID NO: 174 · 248 aa · MW at 29227.4kD
NOV57b, CG59354-02 Protein Sequence	MTTLYDKLLGEKLQYYYSSSEDEDSDHEDKDRGISVNTGPKGVINDWRRFKQLETEQR EEQCREMERLIKKLSMTCRSHLDEEEEQQKQKDLQEKISGKMTLKEFAIMNEDQDDEE FLQQYRKQRMEEMRQQLHKGPQFKQVFEISSGEGFLDMIDKEQKSIVIMVHIYEDGIR DRSHEWLHDPPCKGGELIGNFVRVTDQLGDDFFAVDLEAFLQEFGLLPEKEVLVLTSV RNSATCHSEDSDLEID
	SEQ ID NO: 175 · 891 bp
NOV57c, CG59354-03 DNA Sequence	CACCATGACCACCCTGTATGATAAGTTGCTGGGGGAGAAACTGCAGTACTACTATAGC AGCAGTGAAGATGAGGACAGTGACCACGAGGACAAGGACCGAGGCATCTCAGTTAACA CAGGCCCAAAAGGTGTGATCAATGACTGGCGCCGCTTCAAGCAGTTGGAGACAGAGCA GAGGGAGGAGCAGTGCCGGGAGATGGAAAGGCTGATCAAGAAGCTGTCAATGACTTGC AGGTCCCATCTGGATGAAGAGGAGGAGCAACAGAAACAGAAAGACCTCCAGGAGAAGA TCAGTGGGAAGATGACTCTGAAGGAGTTTGCCATAATGAATGAGGACCAAGATGATGA AGAGTTTCTGCAGCAGTACCGGAAGCAGCGAATGGAAGAGATGCGGCAGCAGCTTCAC AAGGGGCCCCAATTCAAGCAGGTTTTTGAGATCTCCAGTGGAGAAGGGTTTTTAGACA TGATTGATAAAGAACAGAAAAGCATTGTCATCATGGTTCATATTTATGAGGATGGCAT TCCAGGGACCGAAGCCATGAATGGTTGCATGATCCGCCTTGCCGCAGAGTACCCAGCT GTCAAGTTCTGCAAGGTGAAGAGCTCAGTTATTGGCGCCAGCAGTCAGTTCACCAGGA ATGCCCTTCCTGCCCTGCTGATCTATAAGGGGGGTGAATTGATCGGCAATTTTGTTCG TGTTACTGACCAGCTGGGGGATGATTTCTTTGCTGTGGACCTTGAAGCTTTTCTCCAG GAATTTGGATTACTCCCAGAAAAGGAAGTCTTGGTGCTGACATCTGTGCGTAACTCTG CCACGTGTCACAGTGAGGATAGCGACCTGGAAATAGAT**TGA**ACTGATAGTCTAGTTGC ATAGATTTCTCATTGTTTGGG
	ORF Start: ATG at 5 · **ORF Stop: TGA at 851**
	SEQ ID NO: 176 · 282 aa · MW at 32598.5kD
NOV57c, CG59354-03 Protein Sequence	MTTLYDKLLGEKLQYYYSSSEDEDSDHEDKDRGISVNTGPKGVINDWRRFKQLETEQR EEQCREMERLIKKLSMTCRSHLDEEEEQQKQKDLQEKISGKMTLKEFAIMNEDQDDEE FLQQYRKQRMEEMRQQLHKGPQFKQVFEISSGEGFLDMIDKEQKSIVIMVHIYEDGIP GTEAMNGCMIRLAAEYPAVKFCKVKSSVIGASSQFTRNALPALLIYKGGELIGNFVRV TDQLGDDFFAVDLEAFLQEFGLLPEKEVLVLTSVRNSATCHSEDSDLEID

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 57B.

267

Table 57B. Comparison of NOV57a against NOV57b through NOV57c.		
Protein Sequence	NOV57a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV57b	58..240 100..248	138/183 (75%) 140/183 (76%)
NOV57c	58..240 100..282	182/183 (99%) 182/183 (99%)

Further analysis of the NOV57a protein yielded the following properties shown in Table 57C.

Table 57C. Protein Sequence Properties NOV57a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV57a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 57D.

Table 57D. Geneseq Results for NOV57a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV57a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE03537	Human secreted protein variant, SEQ ID NO: 228 - Homo sapiens, 301 aa. [WO200132675-A1, 10-MAY-2001]	1..240 1..301	226/301 (75%) 230/301 (76%)	e-117
AAY99657	Human GTPase associated protein-8 - Homo sapiens, 301 aa. [WO200031263-A2, 02-JUN-2000]	1..240 1..301	226/301 (75%) 230/301 (76%)	e-117
AAE02004	Fruitfly viral IAP-associated factor (VIAF) - Drosophila melanogaster, 240 aa. [WO200134798-A1, 17-MAY-2001]	55..214 59..213	52/161 (32%) 86/161 (53%)	3e-14

268

AAE02003	Zebrafish viral IAP-associated factor (VIAF) - Brachydanio rerio, 239 aa. [WO200134798-A1, 17-MAY-2001]	21..236 2..237	69/241 (28%) 117/241 (47%)	5e-13
AAE02002	Mouse viral IAP-associated factor (VIAF) - Mus musculus, 240 aa. [WO200134798-A1, 17-MAY-2001]	58..240 52..240	59/195 (30%) 99/195 (50%)	4e-12

In a BLAST search of public sequence databases, the NOV57a protein was found to have homology to the proteins shown in the BLASTP data in Table 57E.

Table 57E. Public BLASTP Results for NOV57a				
Protein Accession Number	Protein/Organism/Length	NOV57a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96AF1	HYPOTHETICAL 34.3 KDA PROTEIN - Homo sapiens (Human), 301 aa.	1..240 1..301	226/301 (75%) 230/301 (76%)	e-117
Q13371	Phosducin-like protein (PHLP) - Homo sapiens (Human), 301 aa.	1..240 1..301	225/301 (74%) 230/301 (75%)	e-116
T17321	hypothetical protein DKFZp564M1863.1 - human, 301 aa.	1..240 1..301	225/301 (74%) 230/301 (75%)	e-116
Q923E8	RIKEN CDNA 1200011E13 GENE - Mus musculus (Mouse), 301 aa.	1..240 1..301	210/301 (69%) 223/301 (73%)	e-109
Q63737	Phosducin-like protein (PHLP) - Rattus norvegicus (Rat), 301 aa.	1..240 1..301	210/301 (69%) 223/301 (73%)	e-108

PFam analysis predicts that the NOV57a protein contains the domains shown in the Table 57F.

269

<table>
<tr><td colspan="4" align="center">**Table 57F. Domain Analysis of NOV57a**</td></tr>
</table>

Pfam Domain	NOV57a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Phosducin: domain 1 of 2	35..57	14/23 (61%) 21/23 (91%)	8.7e-08
Phosducin: domain 2 of 2	58..240	133/183 (73%) 174/183 (95%)	9.7e-148

Example 58.

The NOV58 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 58A.

Table 58A. NOV58 Sequence Analysis

	SEQ ID NO: 177	756 bp	
NOV58a, CG59319-01 DNA Sequence			

GGATCCCAATGAAGATACAGAATGGAATGACATTTTAAGAGATTTCGGCATTCTTCCT
CCTAAAGAAGAGTCAAAAGATGAAATTGAAGAAATGGTTTTACGTTTACAGAAAGAAG
CAATGGTGAAACCATTTGAAAAGATGACTCTTGCACAGCTAAAGGAAGCTGAAGATGA
ATTCGATGAAGAAGATATGCAGGCTGTTGAAACATATAGAAAGAAGCGGTTACAGGAA
TGGAAAGCTCTTAAGAAAAAACAAAAATTTGGAGAATTAAGAGAAATTTCTGGAAATC
AGTATGTGAATGAAGTCACAAATGCAGAAGAAGATGTGTGGGTTATAATTCATCTATA
CAGATCAAGCATCCCAATGTGTTTGTTGGTTAACCAGCATCTTAGTCTTCTAGCAAGA
AAGTTTCCAGAAACTAAATTTGTTAAAGCCATCGTGAATAGCTGTATTCAACACTACC
ATGACAATTGTTTACCAACAATTTTTGTGTATAAAAATGGTCAGATAGAAGCCAAATT
CATTGGAATTATAGAATGTGGAGGGATAAATCTCAAGCTGGAAGAACTTGAATGGAAG
CTAGCAGAAGTTGGAGCAATACAGACTGATTTGGAAGAAAACCCCAGAAAAGACATGG
TAGATATGATGGTATCTTCAATTAGAAACACTTCTATTCATGATGACAGTGATAGCTC
CAACAGTGATAATGATACCAAA**TAG**AGAGAAATATTCAATAAATAGCTTTTAGTAAAA
AA

	ORF Start: GAT at 2	ORF Stop: TAG at 719	
	SEQ ID NO: 178	239 aa	MW at 27811.3kD
NOV58a, CG59319-01 Protein Sequence			

DPNEDTEWNDILRDFGILPPKEESKDEIEEMVLRLQKEAMVKPFEKMTLAQLKEAEDE
FDEEDMQAVETYRKKRLQEWKALKKKQKFGELREISGNQYVNEVTNAEEDVWVIIHLY
RSSIPMCLLVNQHLSLLARKFPETKFVKAIVNSCIQHYHDNCLPTIFVYKNGQIEAKF
IGIIECGGINLKLEELEWKLAEVGAIQTDLEENPRKDMVDMMVSSIRNTSIHDDSDSS
NSDNDTK

	SEQ ID NO: 179	745 bp	
NOV58b, CG59319-02 DNA Sequence			

GGATCCCAATGAAGATACAGA**ATG**GATCCCAATGAAGATACAGAATGGAATGACATTT
TAAGAGATTTCGGCATTCTTCCTCCTAAAGAAGAGTCAAAAGATGAAATTGAAGAAAT
GGTTTTACGTTTACAGAAAGAAGCAATGGTGAAACCATTTGAAAAGATGACTCTTGCA
CAGCTAAAGGAAGCTGAAGATGAATTTGATGAAGAAGATATGCAGGCTGTTGAAACAT
ATAGAAAGAAGCGGTTACAGGAATGGAAAGCTCTTAAGAAAAAACAAAAATTTGGAGA
ATTAAGAGAAATTTCTGGAAATCAGTATGTGAATGAAGTCACAAATGCAGAAGAAGAT
GTGTGGGTTATAATTCATCTATACAGATCAAGCATCCCAATGTGTTTGTTGGTTAACC
AGCATCTTAGTCTTCTAGCAAGAAAGTTTCCAGAAACTAAATTTGTTAAAGCCATCGT
GAATAGCTGTATTCAACACTACCATGACAATTGTTTACCAACAATTTTTGTGTATAAA
AATGGTCAGATAGAAGCCAAATTCATTGGAATTATAGAATGTGGAGGGATAAATCTCA
AGCTGGAAGAACTTGAATGGAAGCTAGCAGAAGTTGGAGCAATACAGACTGATTTGGA
AGAAAACCCCAGAAAAGACATGGTAGATATGATGGTATCTTCAATTAGAAACACTTCT
ATCCATGATGACAGTGATAGCTCCAACAGTGATAATGATACCAAA**TAG**A

	ORF Start: ATG at 22	ORF Stop: TAG at 742	
	SEQ ID NO: 180	240 aa	MW at 27942.5kD

270

| NOV58b, CG59319-02 Protein Sequence | MDPNEDTEWNDILRDFGILPPKEESKDEIEEMVLRLQKEAMVKPFEKMTLAQLKEAED EFDEEDMQAVETYRKKRLQEWKALKKKQKFGELREISGNQYVNEVTNAEEDVWVIIHL YRSSIPMCLLVNQHLSLLARKFPETKFVKAIVNSCIQHYHDNCLPTIFVYKNGQIEAK FIGIIECGGINLKLEELEWKLAEVGAIQTDLEENPRKDMVDMMVSSIRNTSIHDDSDS SNSDNDTK |

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 58B.

Table 58B. Comparison of NOV58a against NOV58b.		
Protein Sequence	**NOV58a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV58b	1..239 2..240	216/239 (90%) 216/239 (90%)

Further analysis of the NOV58a protein yielded the following properties shown in Table 58C.

Table 58C. Protein Sequence Properties NOV58a	
PSort analysis:	0.8800 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5   A search of the NOV58a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 58D.

Table 58D. Geneseq Results for NOV58a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV58a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE02003	Zebrafish viral IAP-associated factor (VIAF) - Brachydanio rerio, 239 aa. [WO200134798-A1, 17-MAY-2001]	1..237 3..239	133/238 (55%) 181/238 (75%)	3e-75
AAU27979	Mouse contig polypeptide sequence #132 - Mus musculus, 243 aa. [WO200164834-A2, 07-SEP-2001]	1..231 7..240	137/234 (58%) 176/234 (74%)	2e-74

271

AAU27807	Human full-length polypeptide sequence #132 - Mus musculus, 239 aa. [WO200164834-A2, 07-SEP-2001]	1..231 3..236	137/234 (58%) 176/234 (74%)	2e-74
AAE02001	Human viral IAP-associated factor (VIAF) - Homo sapiens, 239 aa. [WO200134798-A1, 17-MAY-2001]	1..231 3..236	137/234 (58%) 176/234 (74%)	2e-74
AAB68507	Human GTP-binding associated protein #7 - Homo sapiens, 239 aa. [WO200105970-A2, 25-JAN-2001]	1..231 3..236	137/234 (58%) 176/234 (74%)	2e-74

In a BLAST search of public sequence databases, the NOV58a protein was found to have homology to the proteins shown in the BLASTP data in Table 58E.

Table 58E. Public BLASTP Results for NOV58a				
Protein Accession Number	Protein/Organism/Length	NOV58a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9CQU4	1700010B22RIK PROTEIN - Mus musculus (Mouse), 240 aa.	1..239 3..240	208/239 (87%) 229/239 (95%)	e-121
Q9WUP3	PDCL2 - Mus musculus (Mouse), 238 aa (fragment).	1..239 1..238	207/239 (86%) 228/239 (94%)	e-121
Q9DA99	1700016K07RIK PROTEIN - Mus musculus (Mouse), 192 aa.	47..239 1..192	165/193 (85%) 183/193 (94%)	3e-94
CAC40345	SEQUENCE 5 FROM PATENT WO0134798 - Brachydanio rerio (Zebrafish) (Zebra danio), 239 aa.	1..237 3..239	133/238 (55%) 181/238 (75%)	1e-74
Q9H2J4	HTPHLP (UNKNOWN) (PROTEIN FOR MGC:3062) - Homo sapiens (Human), 239 aa.	1..231 3..236	137/234 (58%) 176/234 (74%)	8e-74

PFam analysis predicts that the NOV58a protein contains the domains shown in the

5   Table 58F.

272

Table 58F. Domain Analysis of NOV58a			
Pfam Domain	**NOV58a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Phosducin: domain 1 of 1	60..175	32/120 (27%) 55/120 (46%)	5.8

Example 59.

The NOV59 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 59A.

Table 59A. NOV59 Sequence Analysis			
	SEQ ID NO: 181	981 bp	
NOV59a, CG59576-01 DNA Sequence	GCCACCGCGCCCAGCTGGCTTTTGTTTTTTATCCTTCTGCTCCTCATTTACCTATTCA CCATCATTGGTAGTCTTATGGTGTTCTTTGCCATCAAACTGGATTTCTGCCTGCACAG CTCCTTCTATTTCTTCATCAGTGTCCTCTCCTTCCTAGAGATCTGGTATACCACCATC ACCATCCCCAAGATGTTCTTCAACCTAGCCAGTGAGCAGAAGACCACCTCCCTGGATG GTTGCCTATTGCAGATGTATTTCTTTTACTCCCTCGGCATCACTGAGGTTTGCTTGCT CACCACCAGGGCTATGGACAGATACCTGGCCATCTGTAATCACCTTTGCTACCCCACA GTCACGACACCTCAGCTCTACACTCAGGTGATTCTAGGTTGTTGCATCTGTGGCTTCT TCACGCTGCTCCCTGAGATTGCTTGGATATCCACACTGCCATTTTGTGGTCCAAATCA AATCCACAACATTTTCTGTGACCTTGATCCTATCCTGAATCTAGCATGTGTAGACACT GGCCCAGTTGTTTTAATCAAGGTTGTGGACATTGTACATGCTGTGGAGATCATCACAG CTATAATGCTTGTGACTTTGGCTTACGTCCAAATTATTGCAGTGATCCTAAGAAACTG CTCTGCTGATGGATGCCAAAAGGCATTTTCTACCTATGCTTTCCACCTTGCTATTTTC TTAATCTTTTTTGGAAGTGTAGCCCTGATGTACCTGCTCTTCTCTGCCAAGTACTCCT TTTTCTGGGACACAAACCATCAGCCTAATGTTTGCAGTGCTGTCACCGACAACAATCAT CTGTAGTCTGAGGAATAAAGAGATAAAGGAAGCAATAAAAAAGCACATGTGCCAATCA ATGATATGCACACATCATGTCAAATAAGACCAAATACACACCTCTTAATTACCAAAGA ATATTTATACAAATATTTACATTAATACGTTCAGTGTGTTTGTTGCTGCTGTG		
	ORF Start: GCC at 1	ORF Stop: TAA at 895	
	SEQ ID NO: 182	298 aa	MW at 33780.0kD
NOV59a, CG59576-01 Protein Sequence	ATAPSWLLFFILLLLIYLFTIIGSLMVFFAIKLDFCLHSSFYFFISVLSFLEIWYTTI TIPKMFFNLASEQKTTSLDGCLLQMYFFYSLGITEVCLLTTRAMDRYLAICNHLCYPT VTTPQLYTQVILGCCICGFFTLLPEIAWISTLPFCGPNQIHNIFCDLDPILNLACVDT GPVVLIKVVDIVHAVEIITAIMLVTLAYVQIIAVILRNCSADGCQKAFSTYAFHLAIF LIFFGSVALMYLLFSAKYSFFWDTTISLMFAVLSPTTIICSLRNKEIKEAIKKHMCQS MICTHHVK		

Further analysis of the NOV59a protein yielded the following properties shown in Table 59B.

Table 59B. Protein Sequence Properties NOV59a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 25 and 26

273

A search of the NOV59a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 59C.

Table 59C. Geneseq Results for NOV59a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV59a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG72586	Human OR-like polypeptide query sequence, SEQ ID NO: 2267 - Homo sapiens, 289 aa. [WO200127158-A2, 19-APR-2001]	7..295 1..289	286/289 (98%) 286/289 (98%)	e-167
AAG71784	Human olfactory receptor polypeptide, SEQ ID NO: 1465 - Homo sapiens, 289 aa. [WO200127158-A2, 19-APR-2001]	7..295 1..289	286/289 (98%) 286/289 (98%)	e-167
AAG71785	Human olfactory receptor polypeptide, SEQ ID NO: 1466 - Homo sapiens, 318 aa. [WO200127158-A2, 19-APR-2001]	5..292 20..311	175/293 (59%) 217/293 (73%)	6e-95
AAU24721	Human olfactory receptor AOLFR220 - Homo sapiens, 343 aa. [WO200168805-A2, 20-SEP-2001]	7..283 53..328	170/279 (60%) 212/279 (75%)	4e-94
AAG71808	Human olfactory receptor polypeptide, SEQ ID NO: 1489 - Homo sapiens, 317 aa. [WO200127158-A2, 19-APR-2001]	7..283 29..304	170/279 (60%) 212/279 (75%)	4e-94

In a BLAST search of public sequence databases, the NOV59a protein was found to have homology to the proteins shown in the BLASTP data in Table 59D.

Table 59D. Public BLASTP Results for NOV59a				
Protein Accession Number	Protein/Organism/Length	NOV59a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96R35	OLFACTORY RECEPTOR - Homo sapiens (Human), 216 aa (fragment).	50..267 1..216	107/218 (49%) 146/218 (66%)	7e-55

274

Q9EPG2	M51 OLFACTORY RECEPTOR - Mus musculus (Mouse), 314 aa.	2..285 19..303	115/289 (39%) 172/289 (58%)	2e-52
O95007	Olfactory receptor 6B1 (Olfactory receptor 7-3) (OR7-3) - Homo sapiens (Human), 311 aa.	10..285 28..301	109/279 (39%) 170/279 (60%)	1e-51
Q9QWU6	OLFACTORY RECEPTOR I7 - Mus musculus (Mouse), 327 aa.	1..289 20..314	111/298 (37%) 171/298 (57%)	2e-50
P23270	Olfactory receptor-like protein I7 - Rattus norvegicus (Rat), 327 aa.	1..289 20..314	111/298 (37%) 171/298 (57%)	2e-50

PFam analysis predicts that the NOV59a protein contains the domains shown in the Table 59E.

Table 59E. Domain Analysis of NOV59a			
Pfam Domain	NOV59a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1: domain 1 of 1	37..164	30/134 (22%) 90/134 (67%)	5.4e-13

Example 60.

The NOV60 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 60A.

Table 60A. NOV60 Sequence Analysis			
	SEQ ID NO: 183	1201 bp	
NOV60a, CG59557-01 DNA Sequence	AGGATAACTTTATATGTTGCAAAATGACTCACATAGTATATTTTATTTAACCAGCCTA ATTTCAAGGCTGTTTAGTTGCTTGAAAAGAAGGTTTTTATTTGTTCTTTGCATGTACT TAGAATGCTGACTGTGTTTTATGAGCCAACAAGTGAAACCGCTGAAAATATGGATCCA GAGAATCAGACAATGGTGACTGAGTTTTATTTCTCTGATTTTCCTCAATCTAAGAATG GCAGCCTCTTATTCTTCATTCCTATGCTCTTTATTTATATATTCATTCTTGTTGGAAA TTTCATGATTTTCTTTGCTGTCCGACCGGACCCCCATCTCCATAATCCTATGTACAGT TTTATCAGTGTCTTCTCCTTCCTGGAGATTTGGTACACCACCGTGACTATCCCCAAGA TGCTCTCCAACCTTCTCAGTGAACAGAAAACCATCTCTTTCATAGGTTGCCTCCTGCA GATGTACTTCTTCCACTCACTCGGGGTCACAGAAGCCCTAGTCCTCACAGTGATGGCC ATTGACAGGTGTGTAGCCATCTGCAACCCCCTTCGCTATGCAATCACTATGTCCCCTA GACTGTGCATCCAGCTCTCCACTGGCTCTTGCATTTTTGGCTTCCTCATGTTACTGCC AGAGATTGTGTGCATTTCCACTCTTCCATTCTGTGGCGCCAACCAAATTCATCAACTC TTTTGTGACTTTGAACCTGTGCTGCAGTTAGCCTGCACAGATACGTACATAATTCTGG TTGAAGATGTGATCCGTGCTATTTCCATTCTGACCTCTGTCTCTGTCATCACCCTTTT CTATTTAAGAATCATCACGGTGATCCTGAGGATTCCCTCTGGTGAGAGTCGTCAGAAG GCTTTCTTCACATGTGCAGCCCACATTGCTATTTTCTTGCTGTTTTTGGCAGTGTGT CACTCATGTATCTGCGCTTCTCTGTCACATTCCCACCATTACTGGACAAGGCCATTGC ACTGATGTTTGCTGTCCTTGCCCTACTTTTCAACCCAGTAATCTATAGTCTGAGGAAC AAAGATATGAAAAACGCCACCAAGAAAATCCTCTGTTCTCAAAAGATGTTCAATGCCT CTGGGAGCTAATGGAGTTCACACACACCTCTTCAAAGAAATCTCATCATCTCCTTAAG TTTAAAATGCTAACAAATCAGTTTTTTTAAATTACCATGCA	
	ORF Start: ATG at 121	ORF Stop: TAA at 1111	
	SEQ ID NO: 184	330 aa	MW at 37439.1kD
NOV60a,	MLTVFYEPTSETAENMDPENQTMVTEFYFSDFPQSKNGSLLFFIPMLFIYIFILVGNF MIFFAVRPDPHLHNPMYSFISVFSFLEIWYTTVTIPKMLSNLLSEQKTISFIGCLLQM		

CG59557-01 Protein Sequence	YFFHSLGVTEALVLTVMAIDRCVAICNPLRYAITMSPRLCIQLSTGSCIFGFLMLLPE IVCISTLPFCGANQIHQLFCDFEPVLQLACTDTYIILVEDVIRAISILTSVSVITLFY LRIITVILRIPSGESRQKAFFTCAAHIAIFLLFFGSVSLMYLRFSVTFPPLLDKAIAL MFAVLALLFNPVIYSLRNKDMKNATKKILCSQKMFNASGS

Further analysis of the NOV60a protein yielded the following properties shown in Table 60B.

Table 60B. Protein Sequence Properties NOV60a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 67 and 68

A search of the NOV60a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 60C.

Table 60C. Geneseq Results for NOV60a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV60a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG71807	Human olfactory receptor polypeptide, SEQ ID NO: 1488 - Homo sapiens, 319 aa. [WO200127158-A2, 19-APR-2001]	16..330 1..315	313/315 (99%) 314/315 (99%)	e-180
AAG71803	Human olfactory receptor polypeptide, SEQ ID NO: 1484 - Homo sapiens, 315 aa. [WO200127158-A2, 19-APR-2001]	16..329 1..314	219/314 (69%) 259/314 (81%)	e-129
AAU24658	Human olfactory receptor AOLFR156 - Homo sapiens, 331 aa. [WO200168805-A2, 20-SEP-2001]	9..329 10..330	218/321 (67%) 259/321 (79%)	e-128
AAU24721	Human olfactory receptor AOLFR220 - Homo sapiens, 343 aa. [WO200168805-A2, 20-SEP-2001]	20..329 33..342	196/310 (63%) 234/310 (75%)	e-111
AAG71808	Human olfactory receptor	20..323 9..312	195/304 (64%) 232/304 (76%)	e-111

276

	Homo sapiens, 317 aa. [WO200127158-A2, 19-APR-2001]			

In a BLAST search of public sequence databases, the NOV60a protein was found to have homology to the proteins shown in the BLASTP data in Table 60D.

<table>
<tr><th colspan="5">Table 60D. Public BLASTP Results for NOV60a</th></tr>
<tr><th>Protein
Accession
Number</th><th>Protein/Organism/Length</th><th>NOV60a
Residues/
Match
Residues</th><th>Identities/
Similarities for
the Matched
Portion</th><th>Expect
Value</th></tr>
<tr><td>Q9WU86</td><td>ODORANT RECEPTOR S1 -
Mus musculus (Mouse), 324 aa.</td><td>15..324
8..320</td><td>135/315 (42%)
188/315 (58%)</td><td>4e-67</td></tr>
<tr><td>Q9EPG2</td><td>M51 OLFACTORY RECEPTOR
- Mus musculus (Mouse), 314 aa.</td><td>20..325
5..311</td><td>129/307 (42%)
189/307 (61%)</td><td>4e-65</td></tr>
<tr><td>P23270</td><td>Olfactory receptor-like protein I7
- Rattus norvegicus (Rat), 327 aa.</td><td>24..319
10..310</td><td>126/301 (41%)
182/301 (59%)</td><td>8e-65</td></tr>
<tr><td>Q9QWU6</td><td>OLFACTORY RECEPTOR I7 -
Mus musculus (Mouse), 327 aa.</td><td>16..319
1..310</td><td>128/310 (41%)
184/310 (59%)</td><td>9e-64</td></tr>
<tr><td>O13036</td><td>CHICK OLFACTORY
RECEPTOR 7 - Gallus gallus
(Chicken), 323 aa.</td><td>16..319
1..305</td><td>122/305 (40%)
187/305 (61%)</td><td>1e-63</td></tr>
</table>

PFam analysis predicts that the NOV60a protein contains the domains shown in the Table 60E.

<table>
<tr><th colspan="4">Table 60E. Domain Analysis of NOV60a</th></tr>
<tr><th>Pfam Domain</th><th>NOV60a Match Region</th><th>Identities/
Similarities
for the Matched Region</th><th>Expect Value</th></tr>
<tr><td>7tm_1: domain 1 of 1</td><td>56..304</td><td>45/270 (17%)
172/270 (64%)</td><td>2.4e-21</td></tr>
</table>

Example 61.

The NOV61 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 61A.

10

Table 61A. NOV61 Sequence Analysis		
	SEQ ID NO: 185	1061 bp
NOV61a, CG59555-01 DNA Sequence	CAATCTGGTCCTAAGTGATCTTTTTCTTTTTCACAGGGAAATGGGGGAAAATCAGACA ATGGTCACAGAGTTCCTCCTACTGGGATTTCTCCTGGGCCCAAGGATTCAGATGCTCC TCTTTGGGCTCTTCTCCCTGTTCTATATCTTCACCCTGCTGGGGAACGGGGCCATCCT GGGGCTCATCTCACTGGACTCCAGACTCCACACCCCCATGTACTTCTTCCTCTCACAC CTGGCTGTCGTCGACATCGCCTACACCCGCAACACGGTGCCCCAGATGCTGGCGAACC TCCTGCATCCAGCCAAGCCCATCTCCTTTGCTGGCTGCATGACGCAGACCTTTCTCTG TTTGAGTTTTGGACACAGCGAATGTCTCCTGCTGGTGCTGATGTCCTACGATCGTTAC GTGGCCATCTGCCACCCTCTCCGATACTCCGTCATCATGACCTGGAGAGTCTGCATCA CCCTGGCCGTCACTTCCTGGACGTGTGGCTCCCTCCTGGCTCTGGCCCATGTGGTTCT CATCCTAAGACTGCCCTTCTCTGGGCCTCATGAAATCAACCACTTCTTCTGTGAAATC CTGTCTGTCCTCAGGCTGGCCTGTGCTGACACCTGGCTCAACCAGGTGGTCATCTTTG CAGCCTGCGTGTTCTTCCTGGTGGGGCCACCCAGCCCTGGTGCTTGTCTCCTACTCGCA CATCCTGGCGGCCATCCTGAGGATCCAGTCTGGGGAGGGCCGCAGAAAGGCCTTCTCC ACCTGCTCCTCCCACCTCTGCGTGGTGGGACTCTTCTTTGGCAGTGCCATCATCATGT ACATGGCCCCCAAGTCCCGCCATCCTGAGGAGCAGCAAAAGGTCTTTTTTCTATTTTA CAGTTTTTTCAACCCAACACTTAACCCCCTGATTTACAGCCTGAGGAACGGAGAGGTC AAGGGTGCCCTGAGGAGAGCACTGGGCAAGGAAAGTCATTCCTAACTGGTGTGACATT TGACTCTCCCTCCTCAGTCATCTCCTGGAATCTTGGTACCAAATACCACCTAAGTTCA CTACTCTCTTTATATCA
	ORF Start: ATG at 41	ORF Stop: TAA at 971
	SEQ ID NO: 186	310 aa	MW at 34713.8kD
NOV61a, CG59555-01 Protein Sequence	MGENQTMVTEFLLLGFLLGPRIQMLLFGLFSLFYIFTLLGNGAILGLISLDSRLHTPM YFFLSHLAVVDIAYTRNTVPQMLANLLHPAKPISFAGCMTQTFLCLSFGHSECLLLVL MSYDRYVAICHPLRYSVIMTWRVCITLAVTSWTCGSLLALAHVVLILRLPFSGPHEIN HFFCEILSVLRLACADTWLNQVVIFAACVFFLVGPPSLVLVSYSHILAAILRIQSGEG RRKAFSTCSSHLCVVGLFFGSAIIMYMAPKSRHPEEQQKVFFLFYSFFNPTLNPLIYS LRNGEVKGALRRALGKESHS

Further analysis of the NOV61a protein yielded the following properties shown in Table 61B.

Table 61B. Protein Sequence Properties NOV61a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 43 and 44

A search of the NOV61a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 61C.

Table 61C. Geneseq Results for NOV61a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV61a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM29935	Peptide #3972 encoded by probe for measuring placental gene expression - Homo sapiens, 311 aa. [WO200157272-A2, 09-AUG-2001]	1..310 2..311	310/310 (100%) 310/310 (100%)	0.0
AAM17409	Peptide #3843 encoded by probe for measuring cervical gene expression - Homo sapiens, 311 aa. [WO200157278-A2, 09-AUG-2001]	1..310 2..311	310/310 (100%) 310/310 (100%)	0.0
AAG72949	Human olfactory receptor data exploratorium sequence, SEQ ID NO: 2631 - Homo sapiens, 314 aa. [WO200127158-A2, 19-APR-2001]	1..310 2..311	310/310 (100%) 310/310 (100%)	0.0
AAG72187	Human olfactory receptor polypeptide, SEQ ID NO: 1868 - Homo sapiens, 310 aa. [WO200127158-A2, 19-APR-2001]	1..310 1..310	310/310 (100%) 310/310 (100%)	0.0
AAU04577	Human G-protein coupled receptor like protein, GPCR #11 - Homo sapiens, 308 aa. [WO200153454-A2, 26-JUL-2001]	1..310 1..308	288/310 (92%) 294/310 (93%)	e-165

In a BLAST search of public sequence databases, the NOV61a protein was found to have homology to the proteins shown in the BLASTP data in Table 61D.

Table 61D. Public BLASTP Results for NOV61a				
Protein Accession Number	Protein/Organism/Length	NOV61a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96R46	OLFACTORY RECEPTOR - Homo sapiens (Human), 217 aa (fragment).	67..283 1..217	217/217 (100%) 217/217 (100%)	e-125
O95047	Olfactory receptor 2A4 - Homo sapiens (Human), 310 aa.	1..307 1..307	217/307 (70%) 250/307 (80%)	e-122

279

Q9NQN0	DJ1005H11.1 (7 TRANSMEMBRANE RECEPTOR (RHODOPSIN FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN)) - Homo sapiens (Human), 272 aa (fragment).	39..307 1..269	187/269 (69%) 216/269 (79%)	e-103
Q9Z1V2	OLFACTORY RECEPTOR B12 - Mus musculus (Mouse), 223 aa (fragment).	63..285 1..223	172/223 (77%) 190/223 (85%)	9e-98
O43888	OLFACTORY RECEPTOR - Homo sapiens (Human), 217 aa (fragment).	67..282 1..217	173/217 (79%) 188/217 (85%)	1e-97

PFam analysis predicts that the NOV61a protein contains the domains shown in the Table 61E.

Table 61E. Domain Analysis of NOV61a			
Pfam Domain	NOV61a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1: domain 1 of 1	40..289	65/269 (24%) 188/269 (70%)	1.1e-45

Example 62.

The NOV62 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 62A.

Table 62A. NOV62 Sequence Analysis		
	SEQ ID NO: 187	1201 bp
NOV62a, CG59551-01 DNA Sequence	AGTTGGTTGTAAATAATTCTGCTTATATTACCTACAGAGTAAACATTATAGCATTATC ACTCCAGAATCCTTTGTTTCTATGGTTTCCAGATGTTTCCAATGTCTAGATGTTCCAG CTGCCCATCTCTGAGAAATCCAGCTGTGTCTCACAATGGATGCCACAGCCTGTAATGA ATCAGTGGATGGCTCACCCGTCTTCTATCTATTGGGCATCCCCTCTCTGCCAGAGACC TTCTTCCTCCCTGTGTTTTTTATTTTCCTCCTCTTCTACCTTCTCATCCTGATGGGTA ATGCCCTGATCCTGGTGGCCGTGGTGGCAGAGCCCAGCCTCCACAAGCCCATGTACTT CTTTCTGATCAATCTCTCCACCTTGGACATCCTTTTCACCACAACCACTGTCCCCAAG ATGCTGTCCTTATTCTTGCTTGGGGACCGCTTCCTCAGCTTTTCTTCCTGCTTACTGC AGATGTACCTCTTCCAAAGTTTTACATGTTCAGAAGCCTTCATCCTGGTGGTCATGGC CTATGACCGCTATGTGGCTATCTGCCACCCACTGCACTACCCTGTCCTCATGAACCCA CAGACCAATGCTACCTTGGCAGCCAGTGCCTGGCTAACTGCCCTCCTCCTGCCCATCC CAGCAGTAGTAAGGACCTCCCAGATGGCATATAACAGCATTGCCTACATCTACCACTG CTTCTGTGATCATCTGGCTGTGGTCCAGGCCTCCTGCTCTGACACCACCCCCCAGACC CTCATGGGCTTCTGCATCGCCATGGTGGTGTCCTTCCTCCCCCTTCTCCTGGTGCTTC TCTCCTATGTCCACATCCTGGCCTCAGTGCTTCGCATCAGTTCCCTAGAAGGACGGGC AAAAGCCTTCTCCACCTGCAGCTCCCACCTTCTGGTCGTGGGCACCTACTACTCATCT ATTGCCATAGCCTACGTGGCCTACAGGGCTGACCTGCCCCTTGACTTCCATATCATGG GCAATGTGGTATATGCCATTCTCACACCAATTCTCAACCCCCTCATTTACACGCTGAG AAACAGGGATGTAAAGGCAGCCATCACCAAAATCATGTCTCAAGACCCAGGCTGTGAC AGGAGCATTTGACCTTTAAATGCAGCTAACTCTGCTTCCAGGACACCAAATAACAGTG CTTAGCACAGAGAAAGGACTCAATACATGATAATGAAATAA
	ORF Start: ATG at 152	ORF Stop: TGA at 1112
	SEQ ID NO: 188	320 aa	MW at 35502.6kD

| NOV62a, CG59551-01 Protein Sequence | MDATACNESVDGSPVFYLLGIPSLPETFFLPVFFIFLLFYLLILMGNALILVAVVAEP SLHKPMYFFLINLSTLDILFTTTTVPKMLSLFLLGDRFLSFSSCLLQMYLFQSFTCSE AFILVVMAYDRYVAICHPLHYPVLMNPQTNATLAASAWLTALLLPIPAVVRTSQMAYN SIAYIYHCFCDHLAVVQASCSDTTPQTLMGFCIAMVVSFLPLLLVLLSYVHILASVLR ISSLEGRAKAFSTCSSHLLVVGTYYSSIAIAYVAYRADLPLDFHIMGNVVYAILTPIL NPLIYTLRNRDVKAAITKIMSQDPGCDRSI |

Further analysis of the NOV62a protein yielded the following properties shown in Table 62B.

Table 62B. Protein Sequence Properties NOV62a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	Likely cleavage site between residues 57 and 58

A search of the NOV62a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 62C.

Table 62C. Geneseq Results for NOV62a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV62a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG72119	Human olfactory receptor polypeptide, SEQ ID NO: 1800 - Homo sapiens, 295 aa. [WO200127158-A2, 19-APR-2001]	35..290 2..257	213/256 (83%) 228/256 (88%)	e-119
AAU24639	Human olfactory receptor AOLFR134 - Homo sapiens, 325 aa. [WO200168805-A2, 20-SEP-2001]	16..308 17..308	129/293 (44%) 186/293 (63%)	6e-67
AAG72479	Human OR-like polypeptide query sequence, SEQ ID NO: 2160 - Homo sapiens, 324 aa. [WO200127158-A2, 19-APR-2001]	16..308 17..308	129/293 (44%) 186/293 (63%)	6e-67
AAG71590	Human olfactory receptor polypeptide, SEQ ID NO: 1271 - Homo sapiens, 324 aa. [WO200127158-A2, 19-APR-2001]	16..308 17..308	129/293 (44%) 186/293 (63%)	6e-67
AAG71632	Human olfactory receptor	16..315 13..312	126/300 (42%) 179/300 (59%)	3e-64

281

	Homo sapiens, 316 aa. [WO200127158-A2, 19-APR-2001]			

In a BLAST search of public sequence databases, the NOV62a protein was found to have homology to the proteins shown in the BLASTP data in Table 62D.

Table 62D. Public BLASTP Results for NOV62a				
Protein Accession Number	**Protein/Organism/Length**	**NOV62a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9Z236	OLFACTORY RECEPTOR - Rattus norvegicus (Rat), 221 aa (fragment).	70..289 2..221	187/220 (85%) 202/220 (91%)	e-104
CAB43131	OLFACTORY RECEPTOR - Stenella coeruleoalba (Striped dolphin), 172 aa (fragment).	69..240 1..172	136/172 (79%) 148/172 (85%)	1e-73
Q9EPG2	M51 OLFACTORY RECEPTOR - Mus musculus (Mouse), 314 aa.	16..310 12..305	131/295 (44%) 191/295 (64%)	2e-67
Q9H208	HP4 OLFACTORY RECEPTOR - Homo sapiens (Human), 317 aa (fragment).	16..312 12..308	127/297 (42%) 180/297 (59%)	3e-65
Q920G5	OLFACTORY RECEPTOR P3 - Mus musculus (Mouse), 324 aa.	16..308 19..311	126/295 (42%) 180/295 (60%)	1e-62

PFam analysis predicts that the NOV62a protein contains the domains shown in the Table 62E.

Table 62E. Domain Analysis of NOV62a			
Pfam Domain	**NOV62a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1: domain 1 of 1	46..295	58/268 (22%) 179/268 (67%)	4.6e-38

Example 63.

The NOV63 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 63A.

<table>
<tr><td colspan="3" align="center">**Table 63A. NOV63 Sequence Analysis**</td></tr>
</table>

	SEQ ID NO: 189	1042 bp
NOV63a, CG59540-01 DNA Sequence	GACCTTTCATCACACTCTGGTCATTTACAAACTGTTATTAAGGAATGGGGGACAAGCA GCCCTGGGTCACAGAATTCATCCTGGTTGGATTCCAGCTCAGTGCAGAGATGGAGATC TTTCTCTCTTGCATCTTCTCCCTGTTATATCTCTTCAGTCTACTGAGGAATGGCATGA ACATGGGACTCATCTGTCTGGATCCCAGACTACACACCCCCATATACTTCTTCCTGTC ACACTTGGCCGTCATTGACATATACTATGCTTCCAACAATTTGCTCAACATGCTGGAA AACCTAGTGAAACACAAAAAAACTATCTCGTTCATCTCTTGCATTATGCAGATGGCTT TGTATTTGACTTTTGCTGCTGCAGTGTGCATGATTTTGGTGGTGATGTCCTATGACAG ATTTGTGGCGATCTGCCATCCCCTGCATTACACTGTCATCATGAACTGGAGAGTGTGC ACAGTACTGGCTATTACTTCCTGGGCATGTGGATTTTCCCTGGCCCTCATAAATCTAA TTCTCCTTCTAAGGCTGCCCTTCTGTGGGCCCCAGGAGGTGAACCACTTCTTCGGTGA AATTCTGTCTGTCCTCAAACTGGCCTGTGCAGACACCTGGATTAATGAAATTTTTGTC TTTGCTGGTGGTGTGTTTGTCTTAGTCGGGCCCCTTTCCTTGATGCTGATCTCCTACA TGCGCATCCTCTTGGCCATCCTGAAGATCCAGTCAGGCGAGGGCCACAGAAAGGACTT CTCTACCTGCTCCTCCCACCTCTGTGTGGTGGGGTTCTTCTTTGCCAACGCCATTGTC ATGTACATGGCCCCCAAGTCCCGCCATCCCGAGGAGCAGCAGAAGGTCCTTTCCCTGT TTTGCAGCCTTTGGAATCAGGTGCTGAACCCCCCTCTGATCTACAGCTTGAGGAATGC AGAGGTCAAGAGTGCCCCACAAGAGGGCCACT**GA**AGAAGGAGAGGCTGATGTTACAAT CTCAAAGGCACCACGAGGAGAGGGCCTGCTCCGACAAATGGGGAAGTTGGCTTTTT

	ORF Start: ATG at 45	ORF Stop: TGA at 960	
	SEQ ID NO: 190	305 aa	MW at 34554.8kD
NOV63a, CG59540-01 Protein Sequence	MGDKQPWVTEFILVGFQLSAEMEIFLSCIFSLLYLFSLLRNGMNMGLICLDPRLHTPI YFFLSHLAVIDIYYASNNLLNMLENLVKHKKTISFISCIMQMALYLTFAAAVCMILVV MSYDRFVAICHPLHYTVIMNWRVCTVLAITSWACGFSLALINLILLLRLPFCGPQEVN HFFGEILSVLKLACADTWINEIFVFAGGVFVLVGPLSLMLISYMRILLAILKIQSGEG HRKDFSTCSSHLCVVGFFFANAIVMYMAPKSRHPEEQQKVLSLFCSLWNQVLNPPLIY SLRNAEVKSAPQEGH		

Further analysis of the NOV63a protein yielded the following properties shown in Table 63B.

<table>
<tr><td colspan="2" align="center">**Table 63B. Protein Sequence Properties NOV63a**</td></tr>
</table>

PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	Likely cleavage site between residues 43 and 44

A search of the NOV63a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5 several homologous proteins shown in Table 63C.

<table>
<tr><td colspan="5" align="center">**Table 63C. Geneseq Results for NOV63a**</td></tr>
<tr>
<td>**Geneseq Identifier**</td>
<td>**Protein/Organism/Length [Patent #, Date]**</td>
<td>**NOV63a Residues/ Match Residues**</td>
<td>**Identities/ Similarities for the Matched Region**</td>
<td>**Expect Value**</td>
</tr>
<tr>
<td>AAU24758</td>
<td>Human olfactory receptor</td>
<td>1..300
1..299</td>
<td>258/300 (86%)
275/300 (91%)</td>
<td>e-146</td>
</tr>
</table>

283

	[WO200168805-A2, 20-SEP-2001]			
AAG72952	Human olfactory receptor data exploratorium sequence, SEQ ID NO: 2634 - Homo sapiens, 310 aa. [WO200127158-A2, 19-APR-2001]	1..300 1..299	255/300 (85%) 272/300 (90%)	e-144
AAG72377	Human OR-like polypeptide query sequence, SEQ ID NO: 2058 - Homo sapiens, 312 aa. [WO200127158-A2, 19-APR-2001]	1..300 1..299	255/300 (85%) 272/300 (90%)	e-144
AAG72169	Human olfactory receptor polypeptide, SEQ ID NO: 1850 - Homo sapiens, 312 aa. [WO200127158-A2, 19-APR-2001]	1..300 1..299	255/300 (85%) 272/300 (90%)	e-144
AAG71994	Human olfactory receptor polypeptide, SEQ ID NO: 1675 - Homo sapiens, 314 aa. [WO200127158-A2, 19-APR-2001]	1..300 1..299	225/300 (75%) 256/300 (85%)	e-129

In a BLAST search of public sequence databases, the NOV63a protein was found to have homology to the proteins shown in the BLASTP data in Table 63D.

Table 63D. Public BLASTP Results for NOV63a				
Protein Accession Number	Protein/Organism/Length	NOV63a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O95047	Olfactory receptor 2A4 - Homo sapiens (Human), 310 aa.	1..299 1..298	173/299 (57%) 217/299 (71%)	2e-92
O43885	OLFACTORY RECEPTOR - Homo sapiens (Human), 217 aa (fragment).	67..281 1..216	154/216 (71%) 182/216 (83%)	5e-88
O43888	OLFACTORY RECEPTOR - Homo sapiens (Human), 217 aa (fragment).	67..281 1..216	153/216 (70%) 182/216 (83%)	8e-88
Q96R48	OLFACTORY RECEPTOR - Homo sapiens (Human), 217 aa (fragment).	67..281 1..216	153/216 (70%) 181/216 (82%)	2e-87
Q96R47	OLFACTORY RECEPTOR - Homo sapiens (Human), 215 aa (fragment).	67..281 1..214	149/215 (69%) 175/215 (81%)	3e-84

PFam analysis predicts that the NOV63a protein contains the domains shown in the Table 63E.

Table 63E. Domain Analysis of NOV63a			
Pfam Domain	NOV63a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1: domain 1 of 1	47..290	55/270 (20%) 174/270 (64%)	9.7e-25

Example 64.

The NOV64 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 64A.

Table 64A. NOV64 Sequence Analysis		
	SEQ ID NO: 191	973 bp
NOV64a, CG59280-01 DNA Sequence	AGGCACTAAATGAATATCTGTTTAATTCATAAAGTAACAGAGTTTCTCTTCTCTGGAT TCCCACAGTTTGAAGATGGTAGCCTCCTCTTCTTCATTCCATTGTTTGTTATCTACAT ATTCATTGTCATTGGGAATCTTATTGTATTTTTTGCAGTCAGGGTGGATACCCGTCTC CACAACCCCATGTATAATTTTATCAGCATTTTCTCATTTCTGGAGATCTGGTACACAA CTGCCACAATTCCCAAGATGCTCTCCATCCTCATCAGCAGGCAGAGGACCATCTCCAT GGTTGGTTGCCTCTTGCAGATGTACTTCTTCCATTCACTGGGAAATTCAGAGGGGATT TTGTTGACCACCATGGCCATTGATAGGTACGTTGCCATCTGTAACCCTCTCCGCTACC CAACCATCATGACCCCCGGGCTCTGTGTTCAGCTCTCTGTGGGGTCCTGCATCTTTGG CTTTCTTGTGTTGCTCCCAGAGATTGCATGGATTTCCACACTGCCCTTCTGTGGACCC AACCAAATCCACCAGATCTTCTGTGATTTTGAACCTGTGCTGCGCTTGGCCTGTACAG ACACGTCCATGATTCTGATTGAGGATGTGATCCATGCTGTGGCCATTGTATTCTCTGT CCTGATTATTGCCTTTTCTTATATCAGAATCATCACTGTAATCCTGAGGATTCCCTCT GTTGAAGGCCGCCAGAAGGCCTTTTCTACCTGTGCCGCCCATCTTAGTGTCTTTCTGA TGTTCTATGGCAGTGTATCCCTCATGTACCTGCGTTTCTCTGCCACTTTCCCACCGAT TTTGGACACAGCTGTTGCACTGATGTTTGCAGTTCTTGCTCCCTTTTTCAACCCTATC ATCTATAGCTTTAGAAATAAGGACATGAAGATTGCAATTAAAAAGCTTTTCTGCCCTC AGAAGATGGTTAATTTATCTGTAGATTAATGCTAGCTCATAGGCA
	ORF Start: ATG at 10	ORF Stop: TAA at 955
	SEQ ID NO: 192	315 aa	MW at 35741.4kD
NOV64a, CG59280-01 Protein Sequence	MNICLIHKVTEFLFSGFPQFEDGSLLFFIPLFVIYIFIVIGNLIVFFAVRVDTRLHNP MYNFISIFSFLEIWYTTATIPKMLSILISRQRTISMVGCLLQMYFFHSLGNSEGILLT TMAIDRYVAICNPLRYPTIMTPGLCVQLSVGSCIFGFLVLLPEIAWISTLPFCGPNQI HQIFCDFEPVLRLACTDTSMILIEDVIHAVAIVFSVLIIAFSYIRIITVILRIPSVEG RQKAFSTCAAHLSVFLMFYGSVSLMYLRFSATFPPILDTAVALMFAVLAPFFNPIIYS FRNKDMKIAIKKLFCPQKMVNLSVD	
	SEQ ID NO: 193	929 bp
NOV64b, CG59280-02 DNA Sequence	TCTTCTTCATTCCATTGTTTGTTATCTACATATTCATTGTCATTGGGAATCTTATTGT ATTTTTTGCAGTCAGGGTGGATACCCGTCTCCACAACCCCATGTATAATTTTATCAGC ATTTTCTCATTTCTGGAGATCTGGTACACAACTGCCACAATTCCCAAGATGCTCTCCA TCCTCATCAGCAGGCAGAGGACCATCTCCATGGTTGGTTGCCTCTTGCAGATGTACTT CTTCCATTCACTGGGAAATTCAGAGGGGATTTTGTTGACCACCATGGCCATTGATAGG TACGTTGCCATCTGTAACCCTCTCCGCTACCCAACCATCATGACCCCCGGGCTCTGTG TTCAGCTCTCTGTGGGGTCCTGCATCTTTGGCTTTCTTGTGTTGCTCCCAGAGATTGC ATGGATTTCCACACTGCCCTTCTGTGGACCCAACCAAATCCACCAGATCTTCTGTGAT TTTGAACCTGTGCTGCGCTTGGCCTGTACAGACACGTCCATGATTCTGATTGAGGATG TGATCCATGCTGTGGCCATTGTATTCTCTGTCCTGATTATTGCCTTTTCTTATATCAG AATCATCACTGTAATCCTGAGGATTCCCTCTGTTGAAGGCCGCCAGAAGGCCTTTTCT ACCTGTGCCGCCCATCTTAGTGTCTTTCTGATGTTCTATGGCAGTGTATCCCTCATGT ACCTGCGTTTCTCTGCCACTTTCCCACCGATTTTGGACACAGCTGTTGCACTGATGTT TGCAGTTCTTGCTCCCTTTTTCAACCCTATCATCTATAGCTTTAGAAATAAGGACATG AAGATTGCAATTAAAAAGCTTTTCTGCCCTCAGAAGATGGTTAATTTATCTGTAGATT AATGCTAGCTCATAGGCACCTTTCACTGTGGATGTTACTCTAACACAATAAACCATAT

285

	A
	ORF Start: TTC at 3	ORF Stop: TAA at 870
	SEQ ID NO: 194	289 aa	MW at 32772.9kD
NOV64b, CG59280-02 Protein Sequence	FFIPLFVIYIFIVIGNLIVFFAVRVDTRLHNPMYNFISIFSFLEIWYTTATIPKMLSI LISRQRTISMVGCLLQMYFFHSLGNSEGILLTTMAIDRYVAICNPLRYPTIMTPGLCV QLSVGSCIFGFLVLLPEIAWISTLPFCGPNQIHQIFCDFEPVLRLACTDTSMILIEDV IHAVAIVFSVLIIAFSYIRIITVILRIPSVEGRQKAFSTCAAHLSVFLMFYGSVSLMY LRFSATFPPILDTAVALMFAVLAPFFNPIIYSFRNKDMKIAIKKLFCPQKMVNLSVD

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 64B.

Table 64B. Comparison of NOV64a against NOV64b.		
Protein Sequence	NOV64a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV64b	27..315 1..289	289/289 (100%) 289/289 (100%)

Further analysis of the NOV64a protein yielded the following properties shown in Table 64C.

Table 64C. Protein Sequence Properties NOV64a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	Likely cleavage site between residues 54 and 55

5   A search of the NOV64a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 64D.

Table 64D. Geneseq Results for NOV64a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV64a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG71805	Human olfactory receptor polypeptide, SEQ ID NO: 1486 - Homo sapiens, 256 aa. [WO200127158-A2, 19-APR-2001]	59..314 1..256	255/256 (99%) 255/256 (99%)	e-145

286

AAG71803	Human olfactory receptor polypeptide, SEQ ID NO: 1484 - Homo sapiens, 315 aa. [WO200127158-A2, 19-APR-2001]	9..311 9..311	243/303 (80%) 267/303 (87%)	e-143
AAU24658	Human olfactory receptor AOLFR156 - Homo sapiens, 331 aa. [WO200168805-A2, 20-SEP-2001]	9..311 25..327	240/303 (79%) 264/303 (86%)	e-140
AAG71807	Human olfactory receptor polypeptide, SEQ ID NO: 1488 - Homo sapiens, 319 aa. [WO200127158-A2, 19-APR-2001]	9..313 9..313	222/305 (72%) 259/305 (84%)	e-131
AAU24721	Human olfactory receptor AOLFR220 - Homo sapiens, 343 aa. [WO200168805-A2, 20-SEP-2001]	9..308 37..336	209/300 (69%) 242/300 (80%)	e-119

In a BLAST search of public sequence databases, the NOV64a protein was found to have homology to the proteins shown in the BLASTP data in Table 64E.

Table 64E. Public BLASTP Results for NOV64a				
Protein Accession Number	Protein/Organism/Length	NOV64a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9EPG2	M51 OLFACTORY RECEPTOR - Mus musculus (Mouse), 314 aa.	1..302 4..303	137/303 (45%) 194/303 (63%)	2e-71
Q9EPV0	M50 OLFACTORY RECEPTOR (OLFACTORY RECEPTOR M50) - Mus musculus (Mouse), 316 aa.	6..302 4..301	132/298 (44%) 191/298 (63%)	3e-71
Q9EPG1	M50 OLFACTORY RECEPTOR - Mus musculus (Mouse), 316 aa.	6..302 4..301	130/298 (43%) 190/298 (63%)	2e-70
Q9WU86	ODORANT RECEPTOR S1 - Mus musculus (Mouse), 324 aa.	1..310 12..321	133/313 (42%) 190/313 (60%)	4e-69
Q96KK4	DJ994E9.5 (OLFACTORY RECEPTOR, FAMILY 10, SUBFAMILY C, MEMBER 1 (HS6M1-17)) - Homo sapiens (Human), 306 aa.	9..314 2..306	137/307 (44%) 189/307 (60%)	9e-68

PFam analysis predicts that the NOV64a protein contains the domains shown in the

5   Table 64F.

<table>
<tr><th colspan="4">Table 64F. Domain Analysis of NOV64a</th></tr>
<tr><th>Pfam Domain</th><th>NOV64a Match Region</th><th>Identities/
Similarities
for the Matched Region</th><th>Expect Value</th></tr>
<tr><td>7tm_1: domain 1 of 1</td><td>41..289</td><td>51/269 (19%)
179/269 (67%)</td><td>2.2e-33</td></tr>
</table>

Example 65.

The NOV65 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 65A.

<table>
<tr><th colspan="4">Table 65A. NOV65 Sequence Analysis</th></tr>
<tr><td></td><td>SEQ ID NO: 195</td><td colspan="2">972 bp</td></tr>
<tr><td>NOV65a,
CG59568-01 DNA Sequence</td><td colspan="3">GCATGGTGATCCTGTCCTGGGAAAACCAAACGATGAGAGTGGAATTCGTGCTTCAAGG
ATTCTCTTCCATCAGACAGTTAAATATTTTCCTCTTTATGATAATTTTAGTTTTCTAC
ATCTTAACTGTTTCTGGAAACATCCTCATTGTCCTTCTAGTTTTAGTCAGACATCATC
TCCACACCCCTATGTACTTCCTCCTGGTGAACTTGTCCTGTCTGGAGATCTGGTATAC
CTCTAACATCATCCCCAAAATGTTGCTGATTATCATAGCTGAAGAGAAGACTATCTCT
GTGGCTGGCTGGCTGGCACAATTCTACTTCTTCGGATCCCTGGCTGCCACGGAGTGCC
TCTTGCTCACTGTGATGTCCTATGATCGCTACCTAGCCATCTGCCAGCCTCTTTGCTA
CCGTGTCCTCATGACTGGCCCCCCTTTGCATCAGGCTAGCTGCTGGCTCTTGGTTCTGC
TGCTTCCTCCTTACAGCAATCACCATGGTCTTGCTATGTAGACTAACCTTCTGTGGAC
CCTATGAAACTGATCACTTCTTTTGTGACTTCACCCCTCTGGTTCATCTCTCCTGCAT
GGATACCTCAGTGACTGAGACCATTGCCTTTGCCACCTCTTCTGCAGTAACTCTGATC
CCATTTCTCCTCATTGTAGCCTCCTACTCCTGCGTCCTTTCTGCTATCCTAAGAATCC
CATCTTGCACAGGCCAGAAAAAGGCCTTCTCCACCTGCTCTTCCCACCTCACTGTGGT
CATAGTGTTTTATGGGACACTGATTGCCACATACCTTGTGCCCTCAGCCAACTCATCC
CAACTCTTGTGCAAAGGGTCCTCTCTGCTCTACATCATCCTGACACCCATGTTTAACC
CCATCATTTATAGCCTGAGAAATAGAGACATCCATGAAGCTCTGAAGAAGTGCTTGAG
GAAGAAGTCAGGTGTTTGCCTTAGATAATACGAAAAGGAAAAAA</td></tr>
<tr><td></td><td>ORF Start: ATG at 3</td><td colspan="2">ORF Stop: TAA at 954</td></tr>
<tr><td></td><td>SEQ ID NO: 196</td><td>317 aa</td><td>MW at 35713.4kD</td></tr>
<tr><td>NOV65a,
CG59568-01 Protein Sequence</td><td colspan="3">MVILSWENQTMRVEFVLQGFSSIRQLNIFLFMIILVFYILTVSGNILIVLLVLVRHHL
HTPMYFLLVNLSCLEIWYTSNIIPKMLLIIIAEEKTISVAGWLAQFYFFGSLAATECL
LLTVMSYDRYLAICQPLCYRVLMTGPLCIRLAAGSWFCCFLLTAITMVLLCRLTFCGP
YETDHFFCDFTPLVHLSCMDTSVTETIAFATSSAVTLIPFLLIVASYSCVLSAILRIP
SCTGQKKAFSTCSSHLTVVIVFYGTLIATYLVPSANSSQLLCKGSSLLYIILTPMFNP
IIYSLRNRDIHEALKKCLRKKSGVCLR</td></tr>
</table>

Further analysis of the NOV65a protein yielded the following properties shown in
Table 65B.

<table>
<tr><th colspan="2">Table 65B. Protein Sequence Properties NOV65a</th></tr>
<tr><td>PSort
analysis:</td><td>0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3888 probability located in mitochondrial inner membrane; 0.3030 probability located in mitochondrial intermembrane space</td></tr>
<tr><td>SignalP
analysis:</td><td>Likely cleavage site between residues 45 and 46</td></tr>
</table>

288

A search of the NOV65a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 65C.

Table 65C. Geneseq Results for NOV65a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV65a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG72527	Human OR-like polypeptide query sequence, SEQ ID NO: 2208 - Homo sapiens, 316 aa. [WO200127158-A2, 19-APR-2001]	1..316 1..316	315/316 (99%) 315/316 (99%)	0.0
AAG72231	Human olfactory receptor polypeptide, SEQ ID NO: 1912 - Homo sapiens, 316 aa. [WO200127158-A2, 19-APR-2001]	1..316 1..316	315/316 (99%) 315/316 (99%)	0.0
AAG72084	Human olfactory receptor polypeptide, SEQ ID NO: 1765 - Homo sapiens, 316 aa. [WO200127158-A2, 19-APR-2001]	1..316 1..316	315/316 (99%) 315/316 (99%)	0.0
AAG72700	Murine OR-like polypeptide query sequence, SEQ ID NO: 2382 - Mus musculus, 314 aa. [WO200127158-A2, 19-APR-2001]	1..308 3..308	154/308 (50%) 208/308 (67%)	2e-83
AAG71814	Human olfactory receptor polypeptide, SEQ ID NO: 1495 - Homo sapiens, 317 aa. [WO200127158-A2, 19-APR-2001]	8..311 5..308	142/304 (46%) 208/304 (67%)	7e-79

In a BLAST search of public sequence databases, the NOV65a protein was found to have homology to the proteins shown in the BLASTP data in Table 65D.

Table 65D. Public BLASTP Results for NOV65a				
Protein Accession Number	Protein/Organism/Length	NOV65a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9GZK7	Olfactory receptor 11A1 (Hs6M1-18) - Homo sapiens (Human), 315 aa.	1..308 1..306	147/308 (47%) 202/308 (64%)	4e-77

O13036	CHICK OLFACTORY RECEPTOR 7 - Gallus gallus (Chicken), 323 aa.	7..311 4..308	139/305 (45%) 198/305 (64%)	1e-76
Q9JKA6	OLFACTORY RECEPTOR P2 - Mus musculus (Mouse), 315 aa.	4..313 1..311	143/311 (45%) 194/311 (61%)	1e-75
Q9WU86	ODORANT RECEPTOR S1 - Mus musculus (Mouse), 324 aa.	14..308 21..315	144/295 (48%) 189/295 (63%)	2e-75
Q9UGF6	Olfactory receptor 5V1 (Hs6M1-21) - Homo sapiens (Human), 321 aa.	7..305 4..302	138/299 (46%) 199/299 (66%)	5e-75

PFam analysis predicts that the NOV65a protein contains the domains shown in the Table 65E.

Table 65E. Domain Analysis of NOV65a			
Pfam Domain	**NOV65a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
granulin: domain 1 of 1	144..155	7/13 (54%) 11/13 (85%)	1.7
Trypan_glycop: domain 1 of 1	218..241	6/24 (25%) 21/24 (88%)	7.9
7tm_1: domain 1 of 1	44..293	53/268 (20%) 172/268 (64%)	1.5e-31

Example 66.

The NOV66 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 66A.

Table 66A. NOV66 Sequence Analysis		
	SEQ ID NO: 197	987 bp
NOV66a, CG59224-01 DNA Sequence	CATCTTCCTATGTGTCATGTCTCCTCTTAATGACACAAAAATGGAAGTCCTTAGATTC CTCCTTATCGGGATCACTGGACTGGAGAAAAGTCGCACCTGGATATCCATTCCTTTCT TATCTGTGTACCTTCTTTCTTGGATGGGTAATTTTACCGTCCTCTTTTTTATCAAGAC AGAGCAAAGCCTCCATGAACCTATGTATTATTTGCTTTCCATGCTCTCCATCTCTGAC CTAGGGCTGTCTCTGTCTTCCTTACCCATCACTTTGGGACTATTCCTATTTGATGTCC ATGAAATTCATGCAGCTCCATGCTTTGCCCAGGAATTTTTTATCCATCTGTTTACAGT CAGTGAAGCCTCTGTACTGTCTGTAATGGCATTTGACTGGTATGTGGCAATCCACAGT CCTTTGAGATACAGCACTATCTTAACTAGTCCCAGAGCCATCAAAACAGGGGTTCTTC TGACTTCCAAGAATGTTCTTTTGATCCTTCCACTGCCCTTTCTCTTGCAAAGGCTGAG ATATTGTCATCAAAACCTGCTCTCCCACTCCTATTGTCTCCACCAGGATGTCATGAAG CTGATGTGTTCTGACAACACAGTCAATGTTGTCTACGGACTCTGTGCAGGACTTTCTA CTATGCTGGACTTGGTGTTTATTACCTTCTCCTATATGATTTTAAGGGCTGTACTGGG AATTGCTACCCCCAGACAGCAGTTCAAGGCCCTCAACACGTGCATCTCTCACATCTGT GCTGTGCTTATCTTCTATGTGCCCACGCTGAGTGCTGCCATGCTCCACCAGTTTGCCA GGGATGTGTCTCCTATGATCCACGTCCTCATGGCTGATATTTTTCTGCTGGTGCCACC CCTGTTGAATCCCATCGTGTACTGTGTGAAGACCCACCAAATCCGAGAAAAGGTTGTG

290

	GGGAAACTTTGTCCAAAAGTAAGT**TGA**TCAAAGGAATGAGAAAGGGAATGAATGTATA A
	ORF Start: ATG at 17	ORF Stop: TGA at 953
	SEQ ID NO: 198	312 aa	MW at 35250.7kD
NOV66a, CG59224-01 Protein Sequence	MSPLNDTKMEVLRFLLIGITGLEKSRTWISIPFLSVYLLSWMGNFTVLFFIKTEQSLH EPMYYLLSMLSISDLGLSLSSLPITLGLFLFDVHEIHAAPCFAQEFFIHLFTVSEASV LSVMAFDWYVAIHSPLRYSTILTSPRAIKTGVLLTSKNVLLILPLPFLLQRLRYCHQN LLSHSYCLHQDVMKLMCSDNTVNVVYGLCAGLSTMLDLVFITFSYMILRAVLGIATPR QQFKALNTCISHICAVLIFYVPTLSAAMLHQFARDVSPMIHVLMADIFLLVPPLLNPI VYCVKTHQIREKVVGKLCPKVS

Further analysis of the NOV66a protein yielded the following properties shown in Table 66B.

Table 66B. Protein Sequence Properties NOV66a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.2007 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 50 and 51

A search of the NOV66a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 66C.

Table 66C. Geneseq Results for NOV66a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV66a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG72488	Human OR-like polypeptide query sequence, SEQ ID NO: 2169 - Homo sapiens, 319 aa. [WO200127158-A2, 19-APR-2001]	1..312 1..313	309/313 (98%) 309/313 (98%)	e-176
AAG71557	Human olfactory receptor polypeptide, SEQ ID NO: 1238 - Homo sapiens, 319 aa. [WO200127158-A2, 19-APR-2001]	1..312 1..313	309/313 (98%) 309/313 (98%)	e-176
AAU24573	Human olfactory receptor AOLFR63 - Homo sapiens, 313 aa. [WO200168805-A2, 20-SEP-2001]	1..310 1..311	186/311 (59%) 246/311 (78%)	e-109
AAG71558	Human olfactory receptor	1..310 1..311	185/311 (59%) 245/311 (78%)	e-108

	Homo sapiens, 313 aa. [WO200127158-A2, 19-APR-2001]			
AAU24682	Human olfactory receptor AOLFR181 - Homo sapiens, 312 aa. [WO200168805-A2, 20-SEP-2001]	1..307 1..306	188/308 (61%) 237/308 (76%)	e-106

In a BLAST search of public sequence databases, the NOV66a protein was found to have homology to the proteins shown in the BLASTP data in Table 66D.

Table 66D. Public BLASTP Results for NOV66a				
Protein Accession Number	Protein/Organism/Length	NOV66a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAL35109	PROSTATE-SPECIFIC G PROTEIN-COUPLED RECEPTOR RA1C - Mus musculus (Mouse), 320 aa.	14..304 11..303	141/293 (48%) 199/293 (67%)	2e-77
O88628	Olfactory receptor 51E2 (G-protein coupled receptor RA1c) - Rattus norvegicus (Rat), 320 aa.	14..304 11..303	141/293 (48%) 200/293 (68%)	2e-77
CAC38935	SEQUENCE 9 FROM PATENT WO0131014 - Homo sapiens (Human), 318 aa.	5..304 6..306	145/302 (48%) 206/302 (68%)	2e-77
CAC37756	SEQUENCE 1 FROM PATENT WO0125434 - Homo sapiens (Human), 317 aa.	5..304 5..305	145/302 (48%) 206/302 (68%)	3e-77
Q9H255	Olfactory receptor 51E2 (Prostate specific G-protein coupled receptor) (HPRAJ) - Homo sapiens (Human), 320 aa.	14..304 11..303	139/293 (47%) 198/293 (67%)	2e-76

PFam analysis predicts that the NOV66a protein contains the domains shown in the Table 66E.

Table 66E. Domain Analysis of NOV66a			
Pfam Domain	NOV66a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1: domain 1 of 2	43..151	30/111 (27%) 73/111 (66%)	6.3e-14

292

7tm_1: domain 2 of 2	212..292	16/92 (17%) 52/92 (57%)	0.052

Example 67.

The NOV67 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 67A.

Table 67A. NOV67 Sequence Analysis			
	SEQ ID NO: 199	994 bp	
NOV67a, CG59222-01 DNA Sequence	CACAATGTCTGTCTTCAATAGTTCTGCCTTATACCCTCGCTTCCTCCTAACGGGCCTC TCAGGCCTTGAAAGCAGATATGACTTGATTTCCCTGCCCATCTTCTTGGTTTATGCCA CCTCAATTGCCGGGAACATTAGCATCCTCTTCATTATCAGAACTGAGTCTTCCCTCCA CCAACCGATGTATTACTTTCTGTCAATGCTGGCATTCACTGACCTGGGCCTATCTAAC ACTACCTTACCTACCATGTTCAGTGTCTTCTGGTTCCATGCCCGGGAGATCTCCTTCA ATGCTTGTCTGGTCCAAATGTACTTCATTCATGTTTTCTCGATTATTGAGTCAGCTGT ACTCCTGGCTATGGCCTTTGACTGCTTTATAGCAATCTGTGAACCCTTGCGCTATGCA GCCATCCTAACCAATGATGTAATCATTGGGATTGGGTTGGCAATTGCTGGAAGGGCCT TGGCTCTGGTCTTTCCAGCTTCTTTCCTCTTGAAGAGGCTTCAATATCATGATGTCAA TATTCTGTCCTACCTCTTCTGCCTGCACCAGGACCTCATAAAGACGACTGTATCCAAC TGTCGAGTCAGCAGCATCTATGGCCTCATGGTGGTCATCTGTTCCATGGGACTTGATT CAGTGCTTCTCCTCCTCTCCTATGTCCTCATCCTGGGCACAGTGTTGAGTATAGCCTC CAAGGCAGAGAGAGTGAGAGCCCTCAATACTTGCATCTCCCACATCTGTGCTGTACTC ACCTTCTATACACCAATGATTGGGCTATCTATGATCCATCGCTATGGACAGAATGCTT CCTCAATTGTCCATGTGCTGATGGCCAATGTCTACTTGCTGGTTCCACCTCTCATGAA CCCCGTTGTCTACAGTGTTAAGACCAAGCAGATTCGTGACAGAATCTTCAATAAATTC AAGAAACATGAAGTGTAGATGACAGAGATTCTGAAACATAACTTTCCCTCCATTCCCC ATATATTT	
	ORF Start: ATG at 5	ORF Stop: TAG at 944	
	SEQ ID NO: 200	313 aa	MW at 35044.2kD
NOV67a, CG59222-01 Protein Sequence	MSVFNSSALYPRFLLTGLSGLESRYDLISLPIFLVYATSIAGNISILFIIRTESSLHQ PMYYFLSMLAFTDLGLSNTTLPTMFSVFWFHAREISFNACLVQMYFIHVFSIIESAVL LAMAFDCFIAICEPLRYAAILTNDVIIGIGLAIAGRALALVFPASFLLKRLQYHDVNI LSYLFCLHQDLIKTTVSNCRVSSIYGLMVVICSMGLDSVLLLLSYVLILGTVLSIASK AERVRALNTCISHICAVLTFYTPMIGLSMIHRYGQNASSIVHVLMANVYLLVPPLMNP VVYSVKTKQIRDRIFNKFKKHEV		

Further analysis of the NOV67a protein yielded the following properties shown in Table 67B.

Table 67B. Protein Sequence Properties NOV67a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4047 probability located in mitochondrial inner membrane; 0.4000 probability located in Golgi body; 0.3480 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV67a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 67C.

Table 67C. Geneseq Results for NOV67a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV67a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG72605	Human OR-like polypeptide query sequence, SEQ ID NO: 2286 - Homo sapiens, 318 aa. [WO200127158-A2, 19-APR-2001]	1..309 4..313	295/310 (95%) 298/310 (95%)	e-163
AAG71519	Human olfactory receptor polypeptide, SEQ ID NO: 1200 - Homo sapiens, 318 aa. [WO200127158-A2, 19-APR-2001]	1..309 4..313	295/310 (95%) 298/310 (95%)	e-163
AAU24683	Human olfactory receptor AOLFR182 - Homo sapiens, 314 aa. [WO200168805-A2, 20-SEP-2001]	5..308 9..312	178/304 (58%) 235/304 (76%)	e-102
AAG71715	Human olfactory receptor polypeptide, SEQ ID NO: 1396 - Homo sapiens, 314 aa. [WO200127158-A2, 19-APR-2001]	5..308 9..312	178/304 (58%) 235/304 (76%)	e-102
ABB44526	Human GPCR4a polypeptide SEQ ID NO 11 - Homo sapiens, 315 aa. [WO200174904-A2, 11-OCT-2001]	5..308 6..309	169/304 (55%) 227/304 (74%)	2e-96

In a BLAST search of public sequence databases, the NOV67a protein was found to have homology to the proteins shown in the BLASTP data in Table 67D.

Table 67D. Public BLASTP Results for NOV67a				
Protein Accession Number	Protein/Organism/Length	NOV67a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H344	Olfactory receptor 5I2 (HOR5'beta12) - Homo sapiens (Human), 312 aa.	13..308 12..307	154/296 (52%) 221/296 (74%)	2e-91
Q9H2C8	ODORANT RECEPTOR HOR3'BETA1 - Homo sapiens (Human), 321 aa.	2..308 10..316	160/307 (52%) 216/307 (70%)	5e-89
Q9H343	Olfactory receptor 5I1 (HOR5'beta11) - Homo sapiens (Human), 314 aa.	5..312 5..313	156/309 (50%) 223/309 (71%)	9e-89

294

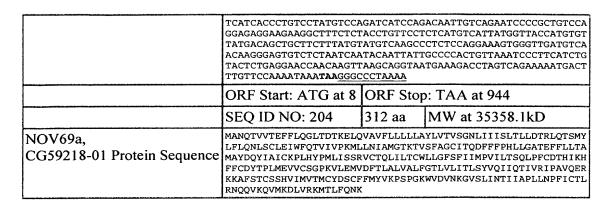| AAL35109	PROSTATE-SPECIFIC G PROTEIN-COUPLED RECEPTOR RA1C - Mus musculus (Mouse), 320 aa.	13..309 11..307	148/297 (49%) 207/297 (68%)	2e-86
Q924X8	OLFACTORY RECEPTOR S85 - Mus musculus (Mouse), 314 aa.	2..304 3..305	150/303 (49%) 221/303 (72%)	1e-85

PFam analysis predicts that the NOV67a protein contains the domains shown in the Table 67E.

Table 67E. Domain Analysis of NOV67a			
Pfam Domain	**NOV67a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1: domain 1 of 1	42..138	24/99 (24%) 67/99 (68%)	7.8e-14

Example 68.

The NOV68 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 68A.

Table 68A. NOV68 Sequence Analysis		
	SEQ ID NO: 201	981 bp
NOV68a, CG59220-01 DNA Sequence	GCAATGAGAAACCGCAGTGTTGTCCCTGAGTTTGTCCTCCTCGGGCTGTCAGCTGGCC CCCAGACCCAGACTCTGCTCTTTGTGCTGTTCGTGGTGATTTGCCTCCTGACTGTGAT GGGAAACCTGCTGCTGCTGGTGGTGATTAATGCTGATTCTTGCCTCCACACACCCATG TACTTCTTCCTGGGACAATTGTCCTTCTTGGATCTCTGCCATTCCTCTGTCACTGCAC CTAAGCTGTTGGAGAACCTCCTGTCTGAGAAGAAAACCATCTCAGTAGAGGGCTGCAT GGCTCAGGTCTTCTTTGTGTTTGCCACTGGGGGCACTGAATCCTCCCTGCTTGCTGTG ATGGCCTATGACCGCTATGTTGCCATCAGCTCTCCTTTGCTCTATGGCCAAGTGATGA ACAGACAGCTGTGTTCAGGGCTGGTGGGGGGCTCATGGGGCTTGGCTTTTCTGGATGC CCTCATCAATATCCTTGTAGCTCTCAATTTAGACTTCTGTGAGGCTCAAAATATCCAC CACTTCAGCTGTGAGCTGCCCTCTCTCTATCCTTTGTCTTGCTCTGATGTGTCAGCAA GTTTTACCACCCTGCTCTGCTCCAGCTTCCTGCATTTCTTTGGAAATTTTCTCATGAT ATTCTTGTCTTATATTTGCATTTTGTCCACCATCCTGAGGATCAGCTCCACTACAGGC AGAAGCAAAGCCTTCTCCACCTGCTCCTCCCACCTCACTGCAGTGATTTTCTTTTATG GCTCCGGATTACTCCGCTATCTCATGCCAAATTCAGGATCCATTCAAGAGCTGATCTT CTCCTTGCAGTACAGCGTGATCACTCCCATGCTGAATCTCCTCATTTACAGCCTGAAG AACAGGGAGGTGAAGGCAGCTGTGAGAAGAACATTGAGAAAATATTTCTAGTGTTTCA ATAGACTTATGAAATCAGAATGATGAGGGAACTGGATAGAACTGCAACAAGCA
	ORF Start: ATG at 4	ORF Stop: TAG at 919
	SEQ ID NO: 202	305 aa	MW at 33732.3kD
NOV68a, CG59220-01 Protein Sequence	MRNRSVVPEFVLLGLSAGPQTQTLLFVLFVVICLLTVMGNLLLLVVINADSCLHTPMY FFLGQLSFLDLCHSSVTAPKLLENLLSEKKTISVEGCMAQVFFVFATGGTESSLLAVM AYDRYVAISSPLLYGQVMNRQLCSGLVGGSWGLAFLDALINILVALNLDFCEAQNIHH FSCELPSLYPLSCSDVSASFTTLLCSSFLHFFGNFLMIFLSYICILSTILRISSTTGR SKAFSTCSSHLTAVIFFYGSGLLRYLMPNSGSIQELIFSLQYSVITPMLNLLIYSLKN REVKAAVRRTLRKYF	

Further analysis of the NOV68a protein yielded the following properties shown in Table 68B.

Table 68B. Protein Sequence Properties NOV68a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 50 and 51

A search of the NOV68a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 68C.

Table 68C. Geneseq Results for NOV68a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV68a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU24771	Human olfactory receptor AOLFR328 - Homo sapiens, 312 aa. [WO200168805-A2, 20-SEP-2001]	3..304 5..306	212/302 (70%) 251/302 (82%)	e-120
AAG98585	Mouse olfactory receptor 7 - Mus musculus domesticus, 214 aa. [WO200146262-A2, 28-JUN-2001]	66..279 1..214	144/214 (67%) 169/214 (78%)	1e-78
AAG72680	Murine OR-like polypeptide query sequence, SEQ ID NO: 2362 - Mus musculus, 337 aa. [WO200127158-A2, 19-APR-2001]	3..305 20..324	148/305 (48%) 201/305 (65%)	3e-74
AAG71546	Human olfactory receptor polypeptide, SEQ ID NO: 1227 - Homo sapiens, 315 aa. [WO200127158-A2, 19-APR-2001]	3..301 5..306	143/302 (47%) 201/302 (66%)	2e-73
AAG66701	Human GPCR1 polypeptide - Homo sapiens, 311 aa. [WO200160865-A2, 23-AUG-2001]	3..301 5..306	143/302 (47%) 201/302 (66%)	2e-73

In a BLAST search of public sequence databases, the NOV68a protein was found to have homology to the proteins shown in the BLASTP data in Table 68D.

Table 68D. Public BLASTP Results for NOV68a

Protein Accession Number	Protein/Organism/Length	NOV68a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9JM36	OLFACTORY RECEPTOR - Mus musculus domesticus (western European house mouse), 214 aa (fragment).	66..279 1..214	144/214 (67%) 169/214 (78%)	5e-78
Q9QZ18	OLFACTORY RECEPTOR - Mus musculus (Mouse), 312 aa.	3..299 5..303	142/299 (47%) 193/299 (64%)	2e-72
Q9EPG6	B1 OLFACTORY RECEPTOR - Mus musculus (Mouse), 314 aa.	3..299 5..303	140/299 (46%) 196/299 (64%)	2e-72
P23266	Olfactory receptor-like protein F5 - Rattus norvegicus (Rat), 313 aa.	3..305 5..309	142/305 (46%) 196/305 (63%)	9e-72
Q9EQA3	ODORANT RECEPTOR K30 - Mus musculus (Mouse), 311 aa.	3..305 5..310	143/306 (46%) 202/306 (65%)	2e-71

PFam analysis predicts that the NOV68a protein contains the domains shown in the Table 68E.

Table 68E. Domain Analysis of NOV68a

Pfam Domain	NOV68a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1: domain 1 of 1	39..286	54/268 (20%) 169/268 (63%)	1.7e-29

Example 69.

The NOV69 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 69A.

Table 69A. NOV69 Sequence Analysis

	SEQ ID NO: 203	957 bp
NOV69a, CG59218-01 DNA Sequence		GTCCACAATGGCCAATCAGACTGTGGTGACTGAGTTCTTCCTCCAAGGCCTGACGGAT ACCAAAGAGCTTCAGGTGGCTGTTTTTCTGCTCCTGCTGCTTGCCTACCTTGTGACTG TCTCTGGGAACCTGATCATCATCAGCCTGACCTTGCTGGACACCCGCCTGCAGACATC TATGTACTTATTTCTCCAGAATCTGTCCTGCTTAGAAATTTGGTTCCAGACAGTCATC GTGCCCAAGATGCTGCTCAACATTGCCATGGGGACCAAGACCGTTAGCTTTGCTGGGT GCATTACCCAGGACTTTTTCTTTCCACATCTTCTGGGGGCCACAGAGTTCTTCCTCCT CACAGCCATGGCCTATGACCAGTATATTGCCATCTGCAAGCCCCTCCACTACCCCATG CTCATAAGTAGTAGAGTCTGCACACAGCTCATCCTCACCTGCTGGCTACTAGGTTTCT CCTTCATCATCATGCCTGTCATCCTGACCAGTCAGCTTCCATTCTGTGATACCCACAT CAAGCATTTCTTCTGTGACTACACGCCTCTAATGGAGGTGGTCTGCAGTGGGCCAAAG GTGCTGGAGATGGTGGATTTTACCCTGGCCTTAGTAGCACTGTTTGGCACCTTGGTAC

297

	TCATCACCCTGTCCTATGTCCAGATCATCCAGACAATTGTCAGAATCCCCGCTGTCCA GGAGAGGAAGAAGGCTTTCTCTACCTGTTCCTCTCATGTCATTATGGTTACCATGTGT TATGACAGCTGCTTCTTTATGTATGTCAAGCCCTCTCCAGGAAAGTGGGTTGATGTCA ACAAGGGAGTGTCTCTAATCAATACAATTATTGCCCCACTGTTAAATCCCTTCATCTG TACTCTGAGGAACCAACAAGTTAAGCAGGTAATGAAAGACCTAGTCAGAAAAATGACT TTGTTCCAAAATAAATAAGGGCCCTAAAA
	ORF Start: ATG at 8 ORF Stop: TAA at 944
	SEQ ID NO: 204 312 aa MW at 35358.1kD
NOV69a, CG59218-01 Protein Sequence	MANQTVVTEFFLQGLTDTKELQVAVFLLLLLAYLVTVSGNLIIISLTLLDTRLQTSMY LFLQNLSCLEIWFQTVIVPKMLLNIAMGTKTVSFAGCITQDFFFPHLLGATEFFLLTA MAYDQYIAICKPLHYPMLISSRVCTQLILTCWLLGFSFIIMPVILTSQLPFCDTHIKH FFCDYTPLMEVVCSGPKVLEMVDFTLALVALFGTLVLITLSYVQIIQTIVRIPAVQER KKAFSTCSSHVIMVTMCYDSCFFMYVKPSPGKWVDVNKGVSLINTIIAPLLNPFICTL RNQQVKQVMKDLVRKMTLFQNK

Further analysis of the NOV69a protein yielded the following properties shown in Table 69B.

Table 69B. Protein Sequence Properties NOV69a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 40 and 41

A search of the NOV69a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 69C.

Table 69C. Geneseq Results for NOV69a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV69a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG72538	Human OR-like polypeptide query sequence, SEQ ID NO: 2219 - Homo sapiens, 313 aa. [WO200127158-A2, 19-APR-2001]	1..312 1..313	284/317 (89%) 293/317 (91%)	e-157
AAG72229	Human olfactory receptor polypeptide, SEQ ID NO: 1910 - Homo sapiens, 313 aa. [WO200127158-A2, 19-APR-2001]	1..312 1..313	284/317 (89%) 293/317 (91%)	e-157
AAU24761	Human olfactory receptor AOLFR112B - Homo sapiens, 309	1..306 1..306	173/307 (56%) 227/307 (73%)	2e-96

	2001]			
AAU24765	Human olfactory receptor AOLFR225B - Homo sapiens, 309 aa. [WO200168805-A2, 20-SEP-2001]	1..306 1..306	166/307 (54%) 227/307 (73%)	2e-94
AAG66353	GPCR partial protein sequence - Unidentified, 313 aa. [WO200155179-A2, 02-AUG-2001]	1..309 1..310	160/311 (51%) 209/311 (66%)	4e-87

In a BLAST search of public sequence databases, the NOV69a protein was found to have homology to the proteins shown in the BLASTP data in Table 69D.

Table 69D. Public BLASTP Results for NOV69a				
Protein Accession Number	Protein/Organism/Length	NOV69a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Z1V0	OLFACTORY RECEPTOR C6 - Mus musculus (Mouse), 313 aa.	1..309 1..310	160/311 (51%) 209/311 (66%)	2e-86
CAC88326	SEQUENCE 18 FROM PATENT WO0164879 - Homo sapiens (Human), 331 aa.	8..306 12..311	142/301 (47%) 200/301 (66%)	4e-78
CAC88328	SEQUENCE 22 FROM PATENT WO0164879 - Homo sapiens (Human), 331 aa.	8..306 12..311	142/301 (47%) 198/301 (65%)	2e-77
CAC88327	SEQUENCE 20 FROM PATENT WO0164879 - Homo sapiens (Human), 331 aa.	8..306 12..311	141/301 (46%) 198/301 (64%)	8e-77
O70270	OLFACTORY RECEPTOR-LIKE PROTEIN - Rattus norvegicus (Rat), 327 aa.	3..308 11..316	136/307 (44%) 208/307 (67%)	4e-76

PFam analysis predicts that the NOV69a protein contains the domains shown in the Table 69E.

Table 69E. Domain Analysis of NOV69a			
Pfam Domain	**NOV69a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1: domain 1 of 1	39..244	47/214 (22%) 147/214 (69%)	1.9e-25

Example 70.

The NOV70 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 70A.

Table 70A. NOV70 Sequence Analysis		
	SEQ ID NO: 205	962 bp
NOV70a, CG59216-01 DNA Sequence	CATCTTCCTATGTGTCATGTCTCCTCTTAATGACACAAAAATGGAAGTCCTTAGATTC CTCCTTATCGGGATCACTGGACTGGAGAAAAGTCGCACCTGGATATCCATTCCTTTCT TATCTGTGTACCTTCTTTCTTGGATGGGTAATTTTACCGTCCTCTTTTTTATCAAGAC AGAGCAAAGCCTCCATGAACCTATGTATTATTTGCTTTCCATGCTCTCCATCTCTGAC CTAGGGCTGTCTCTGTCTTCCTTACCCATCACTTTGGGACTATTCCTATTTGATGTCC ATGAAATTCATGCAGCTCCATGCTTTGCCCAGGAATTTTTTATCCATCTGTTTACAGT CAGTGAAGCCTCTGTACTGTCTGTAATGGCATTTGACTGGTATGTGGCAATCCACAGT CCTTTGAGATACAGCACTATCTTAACTAGTCCCAGAGCCATCAAAACAGGGGTTCTTC TGACTTCCAAGAATGTTCTTTTGATCCTTCCACTGCCCTTTCTCTTGCAAAGGCTGAG ATATTGTCATCAAAACCTGCTCTCCCACTCCTATTGTCTCCACCAGGATGTCATGAAG CTGATGTGTTCTGACAACACAGTCAATGTTGTCTACGGACTCTGTGCAGGACTTTCTA CTATGCTGGACTTGGTGTTTATTACCTTCTCCTATATTATGATTTTAAGGGCTGTACT GGGAATTGCTACCCCCAGACAGCAGTTCAAGGCCCTCAACACGTGCATCTCTCACATC TGTGCTGTGCTTATCTTCTATGTGCCCACGCTGAGTGCTGCCATGCTCCACCAGTTTG CCAGGGATGTGTCTCCTATGATCCACGTCCTCATGGCTGATATTTTTCTGCTGGTGCC ACCCCTGTTGAATCCCATCGTGTACTGTGTGTGAAGACCCACCAAATCCGAGAAAAGGTT GTGGGGAAACTTTGTCCAAAAGTAAGTTGA<u>TCAA</u>

	ORF Start: ATG at 17	ORF Stop: TGA at 956	
	SEQ ID NO: 206	313 aa	MW at 35363.9kD
NOV70a, CG59216-01 Protein Sequence	MSPLNDTKMEVLRFLLIGITGLEKSRTWISIPFLSVYLLSWMGNFTVLFFIKTEQSLH EPMYYLLSMLSISDLGLSLSSLPITLGLFLFDVHEIHAAPCFAQEFFIHLFTVSEASV LSVMAFDWYVAIHSPLRYSTILTSPRAIKTGVLLTSKNVLLILPLPFLLQRLRYCHQN LLSHSYCLHQDVMKLMCSDNTVNVVYGLCAGLSTMLDLVFITFSYIMILRAVLGIATP RQQFKALNTCISHICAVLIFYVPTLSAAMLHQFARDVSPMIHVLMADIFLLVPPLLNP IVYCVKTHQIREKVVGKLCPKVS		

Further analysis of the NOV70a protein yielded the following properties shown in Table 70B.

Table 70B. Protein Sequence Properties NOV70a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.2007 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 50 and 51

A search of the NOV70a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 70C.

Table 70C. Geneseq Results for NOV70a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV70a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG72488	Human OR-like polypeptide query sequence, SEQ ID NO: 2169 - Homo sapiens, 319 aa. [WO200127158-A2, 19-APR-2001]	1..313 1..313	310/313 (99%) 310/313 (99%)	e-178
AAG71557	Human olfactory receptor polypeptide, SEQ ID NO: 1238 - Homo sapiens, 319 aa. [WO200127158-A2, 19-APR-2001]	1..313 1..313	310/313 (99%) 310/313 (99%)	e-178
AAU24573	Human olfactory receptor AOLFR63 - Homo sapiens, 313 aa. [WO200168805-A2, 20-SEP-2001]	1..311 1..311	186/311 (59%) 246/311 (78%)	e-110
AAG71558	Human olfactory receptor polypeptide, SEQ ID NO: 1239 - Homo sapiens, 313 aa. [WO200127158-A2, 19-APR-2001]	1..311 1..311	185/311 (59%) 245/311 (78%)	e-109
AAU24682	Human olfactory receptor AOLFR181 - Homo sapiens, 312 aa. [WO200168805-A2, 20-SEP-2001]	1..308 1..306	188/308 (61%) 238/308 (77%)	e-107

In a BLAST search of public sequence databases, the NOV70a protein was found to have homology to the proteins shown in the BLASTP data in Table 70D.

Table 70D. Public BLASTP Results for NOV70a				
Protein Accession Number	Protein/Organism/Length	NOV70a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC38935	SEQUENCE 9 FROM PATENT WO0131014 - Homo sapiens (Human), 318 aa.	5..305 6..306	145/302 (48%) 207/302 (68%)	5e-79

AAL35109	PROSTATE-SPECIFIC G PROTEIN-COUPLED RECEPTOR RA1C - Mus musculus (Mouse), 320 aa.	14..305 11..303	141/293 (48%) 199/293 (67%)	7e-79
CAC37756	SEQUENCE 1 FROM PATENT WO0125434 - Homo sapiens (Human), 317 aa.	5..305 5..305	145/302 (48%) 207/302 (68%)	7e-79
O88628	Olfactory receptor 51E2 (G-protein coupled receptor RA1c) - Rattus norvegicus (Rat), 320 aa.	14..305 11..303	141/293 (48%) 200/293 (68%)	7e-79
Q9H255	Olfactory receptor 51E2 (Prostate specific G-protein coupled receptor) (HPRAJ) - Homo sapiens (Human), 320 aa.	14..305 11..303	139/293 (47%) 198/293 (67%)	7e-78

PFam analysis predicts that the NOV70a protein contains the domains shown in the Table 70E.

<table>
<tr><th colspan="4">Table 70E. Domain Analysis of NOV70a</th></tr>
<tr><th>Pfam Domain</th><th>NOV70a Match Region</th><th>Identities/ Similarities for the Matched Region</th><th>Expect Value</th></tr>
<tr><td>7tm_1: domain 1 of 2</td><td>43..151</td><td>30/111 (27%) 73/111 (66%)</td><td>6.3e-14</td></tr>
<tr><td>YCF9: domain 1 of 1</td><td>208..262</td><td>10/59 (17%) 31/59 (53%)</td><td>7.5</td></tr>
<tr><td>7tm_1: domain 2 of 2</td><td>212..293</td><td>18/93 (19%) 55/93 (59%)</td><td>0.00034</td></tr>
</table>

Example 71.

The NOV71 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 71A.

<table>
<tr><th colspan="2">Table 71A. NOV71 Sequence Analysis</th></tr>
<tr><td></td><td>SEQ ID NO: 207 | 995 bp</td></tr>
<tr><td>NOV71a,
CG59214-01 DNA Sequence</td><td>GCACAATGTCTGTCTTCAATAGTTCTGCCTTATACCCTCGCTTCCTCCTAACGGGCCT
CTCAGGCCTTGAAAGCAGATATGACTTGATTTCCCTGCCCATCTTCTTGGTTTATGCC
ACCTCAATTGCCGGGAACATTAGCATCCTCTTCATTATCAGAACTGAGTCTTCCCTCC
ACCAACCGATGTATTACTTTCTGTCAATGCTGGCATTCACTGACCTGGGCCTATCTAA
CACTACCTTACCTACCATGTTCAGTGTCTTCTGGTTCCATGCCCGGGAGATCTCCTTC
AATGCTTGTCTGGTCCAAATGTACTTCATTCATGTTTTCTCGATTATTGAGTCAGCTG
TACTCCTGGCTATGGCCTTTGACTGCTTTATAGCAATCTGTGAACCCTTGCGCTATGC
AGCCATCCTAACCAATGATGTAATCATTGGGATTGGGTTGGCAATTGCTGGAAGGGCC
TTGGCTCTGGTCTTTCCAGCTTCTTTCCTCTTGAAGAGGCTTCAATATCATGATGTCA
ATATTCTGTCCTACCTCTTCTGCCTGCACCAGGACCTCATAAAGACGACTGTATCCAA
CTGTCGAGTCAGCAGCATCTATGGCCTCATGGTGGTCATCTGTTCCATGGGACTTGAT
TCAGTGCTTCTCCTCCTCTCCTATGTCCTCATCCTGGGCACAGTGTTGAGTATAGCCT</td></tr>
</table>

302

	CCAAGGCAGAGAGAGTGAGAGCCCTCAATACTTGCATCTCCCACATCTGTGCTGTACT CACCTTCTATACACCAATGATTGGGCTATCTATGATCCATCGCTATGGACAGAATGCT TCCTCAATTGTCCATGTGCTGATGGCCAATGTCTACTTGCTGGTTCCACCTCTCATGA ACCCCGTTGTCTACAGTGTTAAGACCAAGCAGATTCGTGACAGAATCTTCAATAAATT CAAGAAACATGAAGTG**TAG**ATGACAGAGATTCTGAAACATAACTTTCCCTCCATTCCC CATATATTT
	ORF Start: ATG at 6 — ORF Stop: TAG at 945
	SEQ ID NO: 208 — 313 aa — MW at 35044.2kD
NOV71a, CG59214-01 Protein Sequence	MSVFNSSALYPRFLLTGLSGLESRYDLISLPIFLVYATSIAGNISILFIIRTESSLHQ PMYYFLSMLAFTDLGLSNTTLPTMFSVFWFHAREISFNACLVQMYFIHVFSIIESAVL LAMAFDCFIAICEPLRYAAILTNDVIIGIGLAIAGRALALVFPASFLLKRLQYHDVNI LSYLFCLHQDLIKTTVSNCRVSSIYGLMVVICSMGLDSVLLLLSYVLILGTVLSIASK AERVRALNTCISHICAVLTFYTPMIGLSMIHRYGQNASSIVHVLMANVYLLVPPLMNP VVYSVKTKQIRDRIFNKFKKHEV

Further analysis of the NOV71a protein yielded the following properties shown in Table 71B.

Table 71B. Protein Sequence Properties NOV71a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4047 probability located in mitochondrial inner membrane; 0.4000 probability located in Golgi body; 0.3480 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV71a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 71C.

Table 71C. Geneseq Results for NOV71a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV71a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG72605	Human OR-like polypeptide query sequence, SEQ ID NO: 2286 - Homo sapiens, 318 aa. [WO200127158-A2, 19-APR-2001]	1..309 4..313	295/310 (95%) 298/310 (95%)	e-163
AAG71519	Human olfactory receptor polypeptide, SEQ ID NO: 1200 - Homo sapiens, 318 aa. [WO200127158-A2, 19-APR-2001]	1..309 4..313	295/310 (95%) 298/310 (95%)	e-163
AAU24683	Human olfactory receptor	5..308 9..312	178/304 (58%) 235/304 (76%)	e-102

	aa. [WO200168805-A2, 20-SEP-2001]			
AAG71715	Human olfactory receptor polypeptide, SEQ ID NO: 1396 - Homo sapiens, 314 aa. [WO200127158-A2, 19-APR-2001]	5..308 9..312	178/304 (58%) 235/304 (76%)	e-102
ABB44526	Human GPCR4a polypeptide SEQ ID NO 11 - Homo sapiens, 315 aa. [WO200174904-A2, 11-OCT-2001]	5..308 6..309	169/304 (55%) 227/304 (74%)	2e-96

In a BLAST search of public sequence databases, the NOV71a protein was found to have homology to the proteins shown in the BLASTP data in Table 71D.

Table 71D. Public BLASTP Results for NOV71a				
Protein Accession Number	Protein/Organism/Length	NOV71a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H344	Olfactory receptor 5I2 (HOR5'beta12) - Homo sapiens (Human), 312 aa.	13..308 12..307	154/296 (52%) 221/296 (74%)	2e-91
Q9H2C8	ODORANT RECEPTOR HOR3'BETA1 - Homo sapiens (Human), 321 aa.	2..308 10..316	160/307 (52%) 216/307 (70%)	5e-89
Q9H343	Olfactory receptor 5I1 (HOR5'beta11) - Homo sapiens (Human), 314 aa.	5..312 5..313	156/309 (50%) 223/309 (71%)	9e-89
AAL35109	PROSTATE-SPECIFIC G PROTEIN-COUPLED RECEPTOR RA1C - Mus musculus (Mouse), 320 aa.	13..309 11..307	148/297 (49%) 207/297 (68%)	2e-86
Q924X8	OLFACTORY RECEPTOR S85 - Mus musculus (Mouse), 314 aa.	2..304 3..305	150/303 (49%) 221/303 (72%)	1e-85

PFam analysis predicts that the NOV71a protein contains the domains shown in the Table 71E.

Table 71E. Domain Analysis of NOV71a			
Pfam Domain	**NOV71a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1: domain 1 of 1	42..138	24/99 (24%) 67/99 (68%)	7.8e-14

Example 72.

The NOV72 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 72A.

Table 72A. NOV72 Sequence Analysis		
	SEQ ID NO: 209	1004 bp
NOV72a, CG59211-01 DNA Sequence	CTTCTCATCTTTTCCCTCAAATACTGGGA**TG**TCCATTCTCAATACCTCTGAAATGGAA ATCTCTATTTTCTACTTGGTTGGGATCCCAGGTTTGGAGCATGCCAATATTTGGATCT CTATCCCCATATGTCTCATGTACACTGTTGCTATCCTAGGGAATTGTACCATTCTGTT TTTCATAAAAACAGAGCCTTCTTTGCATGAGCCCATGTACTATTTTCTCTCCATGTTG GCTCTCTCTGACCTGGGACTATCCCTCTCCTCTCTCCCTACCATGTTAAGGATTTTCC TGTTCAATGCTCCAGGAATTTCCCCTGATGCCTGTATTGCTCAAGAGTTTTTCATCCA TGGATTCTCAGCTATGGAGTCATCTGTACTTCTTATAATGTCCTTTGATCGCTTTATT GCCATCTGCAACCCCCTGAGATACACTTCCATCCTCACCAGTGCCAGAGTCATTCAAA TTGGGCTTGCTTTTTCTCTCAAAAATGTTTTGTTGATCCTCCCATTTCCTTTCACTCT AAAACATCTAAAATATTGTAAGAAGAACCTCCTGTCCCAATCCTACTGCCTCCATCAA GATGTCATGAAACTGGCCTGCACTGACAACAAGGTCAACATCATCTATGGCTTATTTG TGGCTCTCACAGGCATCCTAGACTTGACATTTATTTTCATGTCCTACATGTTGATACT GAAAGCAGTGTTGAGCATAGCATCAAGAAAGAAAAGGCTCAAGGTCCTCAATACATGT GTTTCCCACATCTGTGCTGTGCTCATCTTCTATGTGCCCATTATCTCCCTAGCTGTCA TCTACCGGTTTGCCAAACACAGTTTCCCAATCACTAGGATCCTCATAGCTGATGCTTT TCTGCTGGTGCCTCCATTGATGAACCCCATTGTATACTGTGTGAAGAGCCAGCAGATA AGAAATCTTGTCTTAGAAAAACTGTGCCAGAAGCAAAGCTGA**AGCGG**ATGCTT**AACCA** CATGATGCTTAACCCAAA
	ORF Start: ATG at 29	ORF Stop: TGA at 968
	SEQ ID NO: 210	313 aa	MW at 35313.1kD
NOV72a, CG59211-01 Protein Sequence	MSILNTSEMEISIFYLVGIPGLEHANIWISIPICLMYTVAILGNCTILFFIKTEPSLH EPMYYFLSMLALSDLGLSLSSLPTMLRIFLFNAPGISPDACIAQEFFIHGFSAMESSV LLIMSFDRFIAICNPLRYTSILTSARVIQIGLAFSLKNVLLILPFPFTLKHLKYCKKN LLSQSYCLHQDVMKLACTDNKVNIIYGLFVALTGILDLTFIFMSYMLILKAVLSIASR KKRLKVLNTCVSHICAVLIFYVPIISLAVIYRFAKHSFPITRILIADAFLLVPPLMNP IVYCVKSQQIRNLVLEKLCQKQS	

Further analysis of the NOV72a protein yielded the following properties shown in Table 72B.

Table 72B. Protein Sequence Properties NOV72a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 44 and 45

305

A search of the NOV72a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 72C.

Table 72C. Geneseq Results for NOV72a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV72a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG71564	Human olfactory receptor polypeptide, SEQ ID NO: 1245 - Homo sapiens, 322 aa. [WO200127158-A2, 19-APR-2001]	1..313 5..317	312/313 (99%) 312/313 (99%)	e-177
AAU24573	Human olfactory receptor AOLFR63 - Homo sapiens, 313 aa. [WO200168805-A2, 20-SEP-2001]	1..312 1..312	225/312 (72%) 272/312 (87%)	e-131
AAG71721	Human olfactory receptor polypeptide, SEQ ID NO: 1402 - Homo sapiens, 316 aa. [WO200127158-A2, 19-APR-2001]	1..311 1..311	236/312 (75%) 267/312 (84%)	e-131
AAU24682	Human olfactory receptor AOLFR181 - Homo sapiens, 312 aa. [WO200168805-A2, 20-SEP-2001]	1..308 1..306	224/308 (72%) 265/308 (85%)	e-131
AAG71701	Human olfactory receptor polypeptide, SEQ ID NO: 1382 - Homo sapiens, 312 aa. [WO200127158-A2, 19-APR-2001]	1..308 1..306	224/308 (72%) 265/308 (85%)	e-131

In a BLAST search of public sequence databases, the NOV72a protein was found to have homology to the proteins shown in the BLASTP data in Table 72D.

Table 72D. Public BLASTP Results for NOV72a				
Protein Accession Number	Protein/Organism/Length	NOV72a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H344	Olfactory receptor 5I2 (HOR5'beta12) - Homo sapiens (Human), 312 aa.	12..304 10..303	152/294 (51%) 219/294 (73%)	6e-90

Q9EQQ7	MOR 3'BETA4 - Mus musculus (Mouse), 319 aa.	1..309 1..310	159/310 (51%) 219/310 (70%)	9e-89
Q9H343	Olfactory receptor 5I11 (HOR5'beta11) - Homo sapiens (Human), 314 aa.	4..313 4..314	154/311 (49%) 226/311 (72%)	9e-89
CAC38935	SEQUENCE 9 FROM PATENT WO0131014 - Homo sapiens (Human), 318 aa.	5..305 6..306	153/302 (50%) 217/302 (71%)	2e-87
CAC37756	SEQUENCE 1 FROM PATENT WO0125434 - Homo sapiens (Human), 317 aa.	5..305 5..305	153/302 (50%) 217/302 (71%)	3e-87

PFam analysis predicts that the NOV72a protein contains the domains shown in the Table 72E.

Table 72E. Domain Analysis of NOV72a			
Pfam Domain	**NOV72a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
DUF40: domain 1 of 1	109..134	10/26 (38%) 20/26 (77%)	0.38
7tm_1: domain 1 of 2	43..144	27/107 (25%) 71/107 (66%)	1.6e-15
7tm_1: domain 2 of 2	212..293	16/93 (17%) 56/93 (60%)	4.7
Sina: domain 1 of 1	300..311	7/12 (58%) 10/12 (83%)	1

Example 73.

The NOV73 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 73A.

Table 73A. NOV73 Sequence Analysis	
SEQ ID NO: 211	1581 bp
NOV73a, CG59276-01 DNA Sequence	CTGGTGGGTTGGCGGCTAAGGGGCGGAGACAAGAGGGGCCGCCACCATCTCCTCCAAT GGAAGGGAGACAGGGGCGGGCTTAATGACGGAAGGAGCATGGCGTGGAGACACCTGAA AAAGCGGGCCCAGGATGCTGTGATCATCCTGGGGGGAGGAGGACTTCTCTTCGCCTCC TACCTGATGGCCACGGGAGATGAGCGTTTCTATGCTGAACACCTGATGCCGACTCTGC AGGGGCTGCTGGACCCGGAGTCAGCCCACAGACTGGCTGTTCGCTTCACCTCCCTGGG GCTCCTTCCACGGGCCAGATTTCAAGACTCTGACATGCTGGAAGTGAGAGTTCTGGGC CATAAATTCCGAAATCCAGTAGGAATTGCTGCAGGATTTGACAAGCATGGGAAGCCG TGGACGGACTTTATAAGATGGGCTTTGGTTTTGTTGAGATAGGAAGTGTGACTCCAAA ACCTCAGGAAGGAAACCCTAGACCCAGAGTCTTCCGCCTCCCTGAGGACCAAGCTGTC ATTAACAGGTATGGATTTAACAGTCACGGGCTTTCAGTGGTGGAACACAGGTTACGGG CCAGACAGCAGAAGCAGGCCAAGCTCACAGAAGATGGACTGCCTCTGGGGGTCAACTT

307

	GGGGAAGAACAAGACCTCAGTGGACGCCGCGGAGGACTACGCAGAAGGGGTGCGCGTA CTGGGCCCCCTGGCCGACTACCTGGTGGTGAATGTGTCCAGCCCCAACÀCTGCCGGGC TGCGGAGCCTTCAGGGAAAGGCCGAGCTGCGCCGCCTGCTGACCAAGGTGCTGCAGGA GAGGGATGGCTTGCGGAGAGTGCÀCAGGCCGGCAGTCCTGGTGAAGATCGCTCCTGAC CTCACCAGCCAGGATAAGGAGGACATTGCCAGTGTGGTCAAAGAGTTGGGCATCGATG GGCTGATTGTTACGAACACCACCGTGAGTCGCCCTGCGGGCCTCCAGGGTGCCCTGCG CTCTGAAACAGGAGGGCTGAGTGGGAAGCCCCTCCGGGATTTATCAACTCAAACCATT CGGGAGATGTATGCACTCACCCAAGGCAAGGTTTCCCGTCGAGTTCCCATAATTGGGG TTGGTGGTGTGAGCAGCGGGCAGGACGCGCTGGAGAAGATCCGGGCAGGGGCCTCCCT GGTGCAGCTGTACACGGCCCTCACCTTCTGGGGGCCACCCGTTGTGGGCAAAGTCAAG CGGGAACTGGAGGCCCTTCTGAAGGAGCAGGGCTTTGGCGGAGTCACAGATGCCATTG GAGCAGATCATCGGAGGATGAGGAAACGGGCAGAGAAGCGGCTGATTGTCCAGTCCCC CTGCGTGGAGGCTGCTTGGCTGGGCTCGAGCCCAGCGGTGGTGGGTCAGTTGGGACCT GGTGGTCTGCTGGTGGTCAGTTTGGGAATTTCCAGGTACGATTGTTTTCAGGCACTGT TCTTTGACTTGGTTGCAGAAAAACAGATTTTGCAACACTTTCCAAGGACACAGTGTTA CCACTCCCTCACCCTGCCATGGCCTCTTGGTTCTGCTTTTAACTTCTGAGCCTCAGGG AGTCCATCTTGTCTG
ORF Start: ATG at 97	ORF Stop: TGA at 1555
SEQ ID NO: 212	486 aa	MW at 52982.6kD

| NOV73a, CG59276-01 Protein Sequence | MAWRHLKKRAQDAVIILGGGGLLFASYLMATGDERFYAEHLMPTLQGLLDPESAHRLA VRFTSLGLLPRARFQDSDMLEVRVLGHKFRNPVGIAAGFDKHGEAVDGLYKMGFGFVE IGSVTPKPQEGNPRPRVFRLPEDQAVINRYGFNSHGLSVVEHRLRARQQKQAKLTEDG LPLGVNLGKNKTSVDAAEDYAEGVRVLGPLADYLVVNVSSPNTAGLRSLQGKAELRRL LTKVLQERDGLRRVHRPAVLVKIAPDLTSQDKEDIASVVKELGIDGLIVTNTTVSRPA GLQGALRSETGGLSGKPLRDLSTQTIREMYALTQGKVSRRVPIIGVGGVSSGQDALEK IRAGASLVQLYTALTFWGPPVVGKVKRELEALLKEQGFGGVTDAIGADHRRMRKRAEK RLIVQSPCVEAAWLGSSPAVVGQLGPGGLLVVSLGISRYDCFQALFFDLVAEKQILQH FPRTQCYHSLTLPWPLGSAFNF |

Further analysis of the NOV73a protein yielded the following properties shown in Table 73B.

Table 73B. Protein Sequence Properties NOV73a	
Psort analysis:	0.8110 probability located in plasma membrane; 0.6400 probability located in endoplasmic reticulum (membrane); 0.3700 probability located in Golgi body; 0.1839 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV73a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 73C.

Table 73C. Geneseq Results for NOV73a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV73a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB70780	Tobacco dihydro-orotase protein - Nicotiana tabacum, 458 aa. [WO200118190-A2, 15-MAR-2001]	36..398 81..458	199/383 (51%) 257/383 (66%)	e-101

308

AAG01301	Human secreted protein, SEQ ID NO: 5382 - Homo sapiens, 144 aa. [EP1033401-A2, 06-SEP-2000]	1..144 1..144	143/144 (99%) 144/144 (99%)	3e-79
AAG91420	C glutamicum protein fragment SEQ ID NO: 5174 - Corynebacterium glutamicum, 371 aa. [EP1108790-A2, 20-JUN-2001]	76..396 60..366	131/328 (39%) 190/328 (56%)	6e-60
AAB46597	C. glutamicum dihydroorotate dehydrogenase protein - Corynebacterium glutamicum, 321 aa. [DE19929364-A1, 28-DEC-2000]	76..396 10..316	131/328 (39%) 190/328 (56%)	6e-60
AAB80123	Corynebacterium glutamicum MP protein sequence SEQ ID NO:980 - Corynebacterium glutamicum, 334 aa. [WO200100843-A2, 04-JAN-2001]	76..396 23..329	131/328 (39%) 190/328 (56%)	1e-59

In a BLAST search of public sequence databases, the NOV73a protein was found to have homology to the proteins shown in the BLASTP data in Table 73D.

Table 73D. Public BLASTP Results for NOV73a				
Protein Accession Number	Protein/Organism/Length	NOV73a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q02127	Dihydroorotate dehydrogenase, mitochondrial precursor (EC 1.3.3.1) (Dihydroorotate oxidase) (DHOdehase) - Homo sapiens (Human), 396 aa (fragment).	1..399 2..396	392/399 (98%) 394/399 (98%)	0.0
PC1219	dihydroorotate oxidase (EC 1.3.3.1) precursor - human, 397 aa.	1..399 3..397	388/399 (97%) 393/399 (98%)	0.0
Q63707	Dihydroorotate dehydrogenase, mitochondrial precursor (EC 1.3.3.1) (Dihydroorotate oxidase) (DHOdehase) - Rattus norvegicus (Rat), 395 aa.	1..399 1..395	350/399 (87%) 369/399 (91%)	0.0
O35435	Dihydroorotate dehydrogenase, mitochondrial precursor (EC 1.3.3.1) (Dihydroorotate oxidase) (DHOdehase) - Mus musculus (Mouse), 395 aa.	1..399 1..395	346/399 (86%) 366/399 (91%)	0.0

309

Q9FZM9	DIHYDROOROTATE DEHYDROGENASE - Oryza sativa (Rice), 468 aa.	29..398 79..468	206/394 (52%) 261/394 (65%)	e-101

PFam analysis predicts that the NOV73a protein contains the domains shown in the Table 73E.

Table 73E. Domain Analysis of NOV73a			
Pfam Domain	NOV73a Match Region	Identities/ Similarities for the Matched Region	Expect Value
DHOdehase: domain 1 of 1	77..381	183/331 (55%) 282/331 (85%)	1.9e-169

Example 74.

The NOV74 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 74A.

Table 74A. NOV74 Sequence Analysis		
	SEQ ID NO: 213	1875 bp
NOV74a, CG59268-01 DNA Sequence	ATGGCCGCAGCCTCGCCTCTGCGCGACTGCCAGGCCTGGAAGGATGCGAGGCTCCCGC TCTCCACCACAAGCAACGAAGCCTGCAAGCTGTTCGATGCCACGCTGACCCAGTATGT AAAATGGACCAATGACAAGAGTCTCGGTGGCATCGAGGGCTGCCTGTCAAAGCTCAAA GCAGCAGATCCAACCTTTGTGATGGGCCACGCCATGGCTACTGGCCTTGTGCTGATTG GCACTGGAAGCTCCGTGAAGCTGGACAAAGAGCTGGACCTGGCTGTGAAGACAATGGT GGAGATTTCAAGAACCCAGCCGCTGACAAGGCGGGAGCAGCTGCACGTGTCTGCAGTA GAGACATTTGCCAATGGGAACTTTCCGAAAGCCTGTGAACTATGGGAACAGATTCTCC AGGACCACCCGACAGACATGTTGGCCCTGAAATTTTCCCATGATGCTTATTTTTACCT GGGCTATCAGGAACAGATGAGAGATTCTGTTGCTCGAATTTACCCCTTCTGGACACCT GACATCCCCCTAAGCAGCTATGTGAAAGGCATCTACTCTTTTGGCTTGATGGAAACCA ACTTCTACGACCAGGCAGAAAACTCGCCAAAGAGGCACCAACTCTTTGTCTTCAACA CCAGCACCCCACAGACAACTACTGGGCAGGAAAAGCAGGCTGTGATGGGGCCAGGAGT GGTAACACATGGGCTCTGTGTCTGCAGCCCCAGGCTGACGCATGGTCGGTGCACACCG TCGCTCACATCCACGAGATGAAAGCAGAGATCAAGGATGGGTTGGAATTCATGCAGCA CTCAGAGACCTTCTGGAAGGACTCTGATATGTTGGCTTGTCATAACTATTGGCACTGG GCTTTATATCTGATTGAGAAGGGTTTAATAAGGAGAACTTTATTCTTCCAGGGCGAAT ATGAGGCCGCGCTGACCATCTACGATACCCACATCCTTCCCAGCCTGCAGGCCAACGA TGCAATGCTGGACGTGGTGGACAGCTGCTCCATGCTCTACCGCCTGCAGATGGAAGGA GTGTCTGTGGGCCAGCGGTGGCAGGATGTCCTGCCTGTGGCCCGGAAGCACAGCCGAG ACCACATCCTGCTGTTCAATGACGCACACTTCCTGATGGCATCCCTGGGTGCACACGA CCCCCAGACCACACAGGAGCTGCTGACCACCCTGCGGGACGCCAGCGAGTATGCAGAG GGGCCTTCTCGGGGTGGGGGTCCTCACCCTGCCGAGAGGTGCCAGGCCTTTGCCTGTA TTATCAGCAATCCTGACGGTTCTGTTAGATTGGCACTGTTATGCCTGCTTACAGATGA GCAAACTGAGGCTGGAAGATCCCCAGGGGAGAACTGCCAGCACCTCCTGGCCCGAGAC GTGGGGCTGCCCCTGTGCCAGGCCCTGGTGGAGGCTGAGGACGGGAACCCTGACCGCG TCCTGGAGCTGCTCCTGCCCATCCGCTACCGGATCGTCCAGCTCGGTGGGAGCAATGC CCAGAGAGACGTCTTCAACCAGCTGCTGATTCACGCGGCCTTAAACTGCACCTCCAGC GTCCATAAGAACGTAGCCCGGAGCCTTCTGATGGAGCGTGATGCCTTGAAGCCCAACT CGCCCCTGACCGAGCGGCTCATCCGCAAGGCAGCTACCGTCCACCTCATGCAGAAGCC TTCTACCCGCCAACCCCCCACTGCAGGCTGCTCTCTCCATGGAAGGAGGCGGCGGCCGC GATGAGCCTTCAGCCTGCCGGGCAGGGGACGTGAACATGGATGACCCTAAGAAGGAAG GCAAGTCCTTGCTGCTGCGGCGCTGTTGTTGTTCAGGATGTTCAGTAGAGATGGAGGG TGATTTAATGTTTCCCTGA
	ORF Start: ATG at 1	ORF Stop: TGA at 1873

310

	SEQ ID NO: 214	624 aa	MW at 69393.3kD
NOV74a, CG59268-01 Protein Sequence	MAAASPLRDCQAWKDARLPLSTTSNEACKLFDATLTQYVKWTNDKSLGGIEGCLSKLK AADPTFVMGHAMATGLVLIGTGSSVKLDKELDLAVKTMVEISRTQPLTRREQLHVSAV ETFANGNFPKACELWEQILQDHPTDMLALKFSHDAYFYLGYQEQMRDSVARIYPFWTP DIPLSSYVKGIYSFGLMETNFYDQAEKLAKEAPTLCLQHQHPTDNYWAGKAGCDGARS GNTWALCLQPQADAWSVHTVAHIHEMKAEIKDGLEFMQHSETFWKDSDMLACHNYWHW ALYLIEKGLIRRTLFFQGEYEAALTIYDTHILPSLQANDAMLDVVDSCSMLYRLQMEG VSVGQRWQDVLPVARKHSRDHILLFNDAHFLMASLGAHDPQTTQELLTTLRDASEYAE GPSRGGGPHPAERCQAFACIISNPDGSVRLALLCLLTDEQTEAGRSPGENCQHLLARD VGLPLCQALVEAEDGNPDRVLELLLPIRYRIVQLGGSNAQRDVFNQLLIHAALNCTSS VHKNVARSLLMERDALKPNSPLTERLIRKAATVHLMQKPSTRQPPLQAALSMEGGGGR DEPSACRAGDVNMDDPKKEGKSLLLRRCCCSGCSVEMEGDLMFP		

Further analysis of the NOV74a protein yielded the following properties shown in Table 74B.

Table 74B. Protein Sequence Properties NOV74a	
PSort analysis:	0.4328 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1137 probability located in mitochondrial inner membrane; 0.1137 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV74a protein against the Geneseq database, a proprietary
5   database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 74C.

Table 74C. Geneseq Results for NOV74a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV74a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41338	Human polypeptide SEQ ID NO 6269 - Homo sapiens, 478 aa. [WO200153312-A1, 26-JUL-2001]	1..559 10..478	463/559 (82%) 466/559 (82%)	0.0
AAM39552	Human polypeptide SEQ ID NO 2697 - Homo sapiens, 453 aa. [WO200153312-A1, 26-JUL-2001]	1..529 1..439	434/529 (82%) 437/529 (82%)	0.0
AAG02871	Human secreted protein, SEQ ID NO: 6952 - Homo sapiens, 104 aa. [EP1033401-A2, 06-SEP-2000]	1..102 1..102	102/102 (100%) 102/102 (100%)	1e-52

AAM40893	Human polypeptide SEQ ID NO 5824 - Homo sapiens, 746 aa. [WO200153312-A1, 26-JUL-2001]	568..604 1..37	32/37 (86%) 32/37 (86%)	2e-10
AAM40892	Human polypeptide SEQ ID NO 5823 - Homo sapiens, 746 aa. [WO200153312-A1, 26-JUL-2001]	568..604 1..37	32/37 (86%) 32/37 (86%)	2e-10

In a BLAST search of public sequence databases, the NOV74a protein was found to have homology to the proteins shown in the BLASTP data in Table 74D.

Table 74D. Public BLASTP Results for NOV74a				
Protein Accession Number	Protein/Organism/Length	NOV74a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH18918	HYPOTHETICAL 45.7 KDA PROTEIN - Homo sapiens (Human), 404 aa.	66..559 1..404	399/494 (80%) 402/494 (80%)	0.0
Q9NWP8	KAIA2372 PROTEIN - Homo sapiens (Human), 336 aa.	1..352 1..310	305/352 (86%) 308/352 (86%)	e-172
Q9XW02	Y54G11A.4 PROTEIN - Caenorhabditis elegans, 497 aa.	4..556 6..458	165/557 (29%) 256/557 (45%)	3e-61
Q9XW01	Y54G11A.7 PROTEIN - Caenorhabditis elegans, 407 aa.	4..347 6..305	122/347 (35%) 177/347 (50%)	7e-53
Q98CS1	MLR5032 PROTEIN - Rhizobium loti (Mesorhizobium loti), 440 aa.	60..553 46..435	145/496 (29%) 215/496 (43%)	1e-43

PFam analysis predicts that the NOV74a protein contains the domains shown in the Table 74E.

Table 74E. Domain Analysis of NOV74a			
Pfam Domain	NOV74a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Monooxygenase: domain 1 of 1	225..410	28/238 (12%) 121/238 (51%)	6.4

312

Example 75.

The NOV75 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 75A.

Table 75A. NOV75 Sequence Analysis		
	SEQ ID NO: 215	1851 bp
NOV75a, CG59549-01 DNA Sequence	CAGCTACAGCAAACATCGTTCGAG**ATG**TCCCACCAAGAGGGCAGCACAGGTGGCTTAC CAGACTTAGTGACTGAAAGCCTGTTCAGCAGCCCAGAGGAGCAGTCTGGAGTAGCAGC GGTGACGGCGGCCTCCTCAGACATTGAAATGGCAGCCACAGAGCCATCGACCGGAGAT GGTGGTGATACCAGGGATGGTGGTTTCCTGAACGATGCCAGCACAGAAAATCAAAACA CAGACTCAGAAAGTTCAAGTGAAGACGTCGAACTTGAAAGCATGGGTGAAGGTTTATT TGGTTACCCGTTAGTGGGAGAGGAGACAGAAAGGGAGGAGGAAGAAGAAGAGATGGAG GAGGAAGGGGAGGAGGAAGAACAGCCTCGGATGTGTCCACGATGCGGTGGCACCAACC ATGATCAGTGTTTGTTAGACGAGGATCAGGCGTTGGAGGAGTGGATTTCCTCAGAGAC ATCTGCCCTGCCCCGATCTCGCTGGCAAGTCCTTACTGCTCTTCGCCAGCGGCAGCTG GGTTCAAGTGCCCGCTTTGTATATGAGGCCTGTGGGGCAAGAACCTTTGTGCAGCGTT TCCGCCTGCAGTATCTTCTTGGAAGCCATGCCGGTTCTGTCAGTACCATACACTTTAA CCAGCGTGGCACCCGACTGGCCAGTAGCGGTGATGACTTAAGGGTGATAGTGTGGGAC TGGGTGCGGCAGAAGCCAGTACTGAACTTTGAGAGTGGTCACGATATTAATGTCATCC AGGCTAAGTTCTTTCCTAACTGTGGTGATTCCACTCTGGCCATGTGTGGCCATGATGG ACAGGTACGGGTAGCAGAACTAATTAATGCATCATATTGCGAGAATACTAAGCGTGTG GCCAAGCACAGGGGACCTGCCCACGAGTTGGCTCTGGAGCCAGACTCTCCTTATAAGT TCCTCACTTCAGGTGAAGATGCCGTTGTGTTCACCATTGACCTCAGGCAAGACCGGCC AGCTTCAAAAGTTGTGGTAACAAGAGAAATGATAAGAAAGTCGGACTGTATACAATC TCTATGAATCCTGCCAATATTTACCAATTTGCAGTGGGTGGACATGATCAGTTTGTAA GGATTTATGACCAGAGGAGAATTGATAAGAAAGAAAACAATGGAGTACTCAAGAAATT CACTCCTCATCATCTCGGTTTATTGTGATTTCCCAACAAACATCACCTGCGTTGTGTAC AGCCACGATGGCACAGAGCTCCTGGCCAGCTACAATGATGAAGATATTTACCTCTTCA ACTCCTCTCTCAGTGATGGTGCTCAATATGTTAAGAGATATAAGGGGCACAGAAATAA TGACACAATCAAATGTGTTAATTTCTATGGCCCCCCGGAGTGAGTTTGTCGTGAGCGGT AGTGATTGTGGGCACGTCTTCTTCTTCTGGGAGAAATCATCCTCCCAGATCATCCAGTTCA TGGAGGGGGACAGAGGAGATATAGTAAACTGTCTTGAACCCCACCCTTACCTACCTGT GTTGGCGACCAGTGGCCTAGATCAGCATGTCAGGATCTGGACACCCACAGCTAAAACT GCCACTGAGCTTACTGGGTTAAAAGATGTGATTAAGAAGAACAAGCAGGAGCGAGATG AAGACAACTTGAACTATACGGACTCGTTTGACAACCGCATGCTTCGGTTCTTCGTGCG TCACCTGTTACAGAGAGCTCATCAACCCGGCTGGAGAGATCATGGAGCTGAGTTCCCA GATGAAGAAGAGTTGGATGAGTCTTCCAGCACCTCAGATACATCCGAGGAGGAGGGCC AAGATCGAGTGCAGTGCATACCATCC**TGA**AGGCCTCATATCCAGTCCAGCTAG
	ORF Start: ATG at 25	ORF Stop: TGA at 1825
	SEQ ID NO: 216	600 aa	MW at 67372.4kD
NOV75a, CG59549-01 Protein Sequence	MSHQEGSTGGLPDLVTESLFSSPEEQSGVAAVTAASSDIEMAATEPSTGDGGDTRDGG FLNDASTENQNTDSESSSEDVELESMGEGLFGYPLVGEETEREEEEEEMEEEGEEEEQ PRMCPRCGGTNHDQCLLDEDQALEEWISSETSALPRSRWQVLTALRQRQLGSSARFVY EACGARTFVQRFRLQYLLGSHAGSVSTIHFNQRGTRLASSGDDLRVIVWDWVRQKPVL NFESGHDINVIQAKFFPNCGDSTLAMCGHDGQVRVAELINASYCENTKRVAKHRGPAH ELALEPDSPYKFLTSGEDAVVFTIDLRQDRPASKVVVTRENDKKVGLYTISMNPANIY QFAVGGHDQFVRIYDQRRIDKKENNGVLKKFTPHHLVYCDFPTNITCVVYSHDGTELL ASYNDEDIYLFNSSLSDGAQYVKRYKGHRNNDTIKCVNFYGPRSEFVVSGSDCGHVFF WEKSSSQIIQFMEGDRGDIVNCLEPHPYLPVLATSGLDQHVRIWTPTAKTATELTGLK DVIKKNKQERDEDNLNYTDSFDNRMLRFFVRHLLQRAHQPGWRDHGAEFPDEEELDES SSTSDTSEEEGQDRVQCIPS

Further analysis of the NOV75a protein yielded the following properties shown in
5   Table 75B.

Table 75B. Protein Sequence Properties NOV75a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0442 probability located in microbody (peroxisome)

313

SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV75a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 75C.

Table 75C. Geneseq Results for NOV75a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV75a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR85870	WD-40 domain-contg. Mus musculus protein - Mus musculus, 816 aa. [WO9521252-A2, 10-AUG-1995]	95..589 333..815	295/495 (59%) 372/495 (74%)	e-179
AAM73935	Human bone marrow expressed probe encoded protein SEQ ID NO: 34241 - Homo sapiens, 164 aa. [WO200157276-A2, 09-AUG-2001]	1..157 8..164	157/157 (100%) 157/157 (100%)	2e-87
AAM61216	Human brain expressed single exon probe encoded protein SEQ ID NO: 33321 - Homo sapiens, 164 aa. [WO200157275-A2, 09-AUG-2001]	1..157 8..164	157/157 (100%) 157/157 (100%)	2e-87
AAM34114	Peptide #8151 encoded by probe for measuring placental gene expression - Homo sapiens, 164 aa. [WO200157272-A2, 09-AUG-2001]	1..157 8..164	157/157 (100%) 157/157 (100%)	2e-87
AAB57007	Human prostate cancer antigen protein sequence SEQ ID NO:1585 - Homo sapiens, 214 aa. [WO200055174-A1, 21-SEP-2000]	408..600 22..214	144/194 (74%) 162/194 (83%)	2e-80

In a BLAST search of public sequence databases, the NOV75a protein was found to
5   have homology to the proteins shown in the BLASTP data in Table 75D.

Table 75D. Public BLASTP Results for NOV75a				
Protein Accession Number	Protein/Organism/Length	NOV75a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q12839	H326 PROTEIN - Homo sapiens (Human), 597 aa.	1..600 1..597	408/604 (67%) 471/604 (77%)	0.0
Q01078	PROTEIN PC326 - Mus musculus (Mouse), 747 aa.	95..589 264..746	295/495 (59%) 372/495 (74%)	e-178
Q9W091	CG8001 PROTEIN - Drosophila melanogaster (Fruit fly), 748 aa.	68..587 209..711	178/533 (33%) 280/533 (52%)	1e-77
Q96E00	UNKNOWN (PROTEIN FOR MGC:9478) - Homo sapiens (Human), 273 aa.	1..246 1..243	141/249 (56%) 173/249 (68%)	8e-66
Q9M1E5	HYPOTHETICAL 54.0 KDA PROTEIN - Arabidopsis thaliana (Mouse-ear cress), 481 aa.	183..536 42..419	136/382 (35%) 209/382 (54%)	2e-62

PFam analysis predicts that the NOV75a protein contains the domains shown in the Table 75E.

Table 75E. Domain Analysis of NOV75a			
Pfam Domain	NOV75a Match Region	Identities/ Similarities for the Matched Region	Expect Value
WD40: domain 1 of 7	188..224	13/37 (35%) 29/37 (78%)	0.0016
WD40: domain 2 of 7	231..269	12/39 (31%) 26/39 (67%)	11
WD40: domain 3 of 7	278..315	9/38 (24%) 24/38 (63%)	2.2e+02
WD40: domain 4 of 7	326..363	8/38 (21%) 27/38 (71%)	8.8
WD40: domain 5 of 7	382..418	5/37 (14%) 27/37 (73%)	12
WD40: domain 6 of 7	429..466	6/38 (16%) 26/38 (68%)	18
WD40: domain 7 of 7	473..509	11/37 (30%) 22/37 (59%)	0.51

315

Example 76.

The NOV76 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 76A.

Table 76A. NOV76 Sequence Analysis		
	SEQ ID NO: 217	7497 bp
NOV76a, CG59641-01 DNA Sequence	ATGGTCTTGCTTCTTTGTCTATCTTGTCTGATTTTCTCCTGTCTGACCTTTTCCTGGT TAAAAATCTGGGGGAAAATGACGGACTCCAAGCCGATCACCAAGAGTAAATCAGAAGC AAACCTCATCCCGAGCCAGGAGCCCTTTCCAGCCTCTGATAACTCAGGGGAGACACCG CAGAGAAATGGGGAGGGCCACACTCTGCCCAAGACACCCAGCCAGGCCGAGCCAGCCT CCCACAAAGGCCCCAAAGATGCCGGTCGGCGGAGAAACTCCCTACCACCCTCCCACCA GAAGCCCCCAAGAAACCCCCTTTCTTCCAGTGACGCAGCACCCTCCCCAGAGCTTCAA GCCAACGGGACTGGGACACAAGGTCTGGAGGCCACAGATACCAATGGCCTGTCCTCCT CAGCCAGGCCCCAGGGCCAGCAAGCTGGCTCCCCCTCCAAAGAAGACAAGAAGCAGGC AAACATCAAGAGGCAGCTGATGACCAACTTCATCCTGGGCTCTTTTGATGACTACTCC TCCGACGAGGACTCTGTTGCTGGCTCATCTCGTGAGTCTACCCGGAAGGGCAGCCGGG CCAGCTTGGGGGCCCTGTCCCTGGAGGCTTATCTGACCACAAGGCCGAGCATGTCGGG ACTCCACCTGGTGAAGAGGGGACGGGAACACAAGAAGCTGGACCTGCACAGAGACTTT ACCGTGGCTTCTCCCGCTGAGTTTGTCACACGCTTTGGGGGGGATCGGGTCATCGAGA AGGTGCTTATTGCCAACAACGGGATTGCCGCCGTGAAGTGCATGCGCTCCATCCGCAG GTGGGCCTATGAGATGTTCCGCAACGAGCGGGCCATCCGGTTTGTTGTGATGGTGACC CCCGAGGACCTTAAGGCCAACGCCAGAGTACATCAAGATGGCGGATCATTACGTCCCCG TCCCAGGAGGGCCCAATAACAACAACTATGCCAACGTGGAGCTGATTGTGGACATTGC CAAGAGAATCCCCGTGCAGGCGGTGTGGGCTGGCTGGGGCCATGCTTCAGAAAACCCT AAACTTCCGGAGCTGCTGTGCAAGAATGGAGTTGCTTTCTTAGGCCCTCCCAGTGAGG CCATGTGGGCCTTAGGAGATAAGATCGCCTCCACCGTTGTCGCCCAGACGCTACAGGT CCCAACCCTGCCCTGGAGTGGAAGCGGCCTGACAGTGGAGTGGACAGAAGATGATCTG CAGCAGGGAAAAAGAATCAGTGTCCCAGAAGATGTTTATGACAAGGGTTGCGTGAAAG ACGTAGATGAGGGCTTGGAGGCAGCAGAAAGAATTGGTTTTCCATTGATGATCAAAGC TTCTGAAGGTGGCGGAGGGAAGGGAATCCGGAAGGCTGAGAGTGCGGAGGACTTCCCG ATCCTTTTCAGACAAGTACAGAGTGAGATCCCAGGCTCGCCCATCTTTCTCATGAAGC TGGCCCAGCACGCCCGTCACCTGGAAGTTCAGATCCTCGCTGACCAGTATGGGAATGC TGTGTCTCTGTTTGGTCGCGACTGCTCCATCCAGCGGCGGCATCAGAAGATCGTTGAG GAAGCACCGGCCACCATCGCCCCGCTGGCCATATTCGAGTTCATGGAGCAGTGTGCCA TCCGCCTGGCCAAGACCGTGGGCTATGTGAGTGCAGGGACAGTGGAATACCTCTATAG TCAGGATGGCAGCTTCCACTTCTTGGAGCTGAATCCTCGCTTGCAGGTGGAACATCCC TGCACAGAAATGATTGCTGATGTTAATCTGCCGGCCGCCCAGCTACAGATCGCCATGG GCGTGCCACTGCACCGGCTGAAGGATATCCGGCTTCTGTATGGAGAGTCACCATGGGG AGTGACTCCCATTTCTTTTGAAACCCCCTCAAACCCTCCCCTCGCCCGAGGCCCACGTC ATTGCCGCCAGAATCACCAGCGAAAACCCAGACGAGGGTTTTAAGCCGAGCTCCGGGA CTGTCCAGGAACTGAATTTCCGGAGCAGCAAGAACGTGTGGGGTTACTTCAGCGTGGC CGCTACTGGAGGCCTGCACGAGTTTGCGGATTCCCAATTTGGGCACTGCTTCTCCTGG GGAGAGAACCGGGAAGAGGCCATTTCGAACATGGTGGTGGCTTTGAAGGAACTGTCCA TCCGAGGCGACTTTAGGACTACCGTGGAATACCTCATTAACCTCCTGGAGACCGAGAG CTTCCAGAACAACGACATCGACACCGGGTGGTTGGACTACCTCATTGCTGAGAAAGTG CAGGCGGAGAAACCGGATATCATGCTTGGGGTGGTATGCGGGGCCTTGAACGTGGCCG ATGCGATGTTCAGAACGTGCATGACAGATTTCTTACACTCCCTGGAAAGGGGGCCAGGT CCTCCCAGCGGATTCACTACTGAACCTCGTAGATGTGGAATTAATTTACGGAGGTGTT AAGTACATTCTCAAGGTGGCCCGGCAGTCTCTGACCATGTTCGTTCTCATCATGAATG GCTGCCACATCGAGATTGATGCCCACCGGCTGAATGATGGGGGGGCTCCTGCTCTCCTA CAATGGGAACAGCTACACCACCTACATGAAGGAAGAGGTTGACAGTTACCGAATTACC ATCGGCAATAAGACGTGTGTGTTTGAGAAGGAGAACGATCCTACAGTCCTGAGATCCC CCTCGGCTGGGAAGCTGACACAGTACACAGTGGAGGATGGGGGCCACGTTGAGGCTGG GAGCAGCTACGCTGAGATGGAGGTGATGAAGATGATCATGACCCTGAACGTTCAGGAA AGAGGCCGGGTGAAGTACATCAAGCGTCCAGGTGCCGTGCTGGAAGCAGGCTGCGTGG TGGCCAGGCTGGAGCTCGATGACCCTTCTAAAGTCCACCCGGCTGAACCGTTCACAGG AGAACTCCCTGCCCAGCAGACACTGCCCATCCTCGGAGAGAAACTGCACCAGGTCTTC CACAGCGTCCTGGAAAACCTCACCAACGTCATGAGTGGCTTTTGTCTGCCAGAGCCCG TTTTTAGCATAAAGCTGAAGGAGTGGGTGCAGAAGCTCATGATGACCCTCCGGCACCC GTCACTGCCGCTGCTGGAGCTGCAGGAGATCATGACCAGCGTGGCAGGCCGCATCCCC GCCCCTGTGGAGAAGTCTGTCCGCAGGGTGATGGCCCAGTATGCCAGCAACATCACCT CGGTGCTGTGCCAGTTCCCCAGCCAGCAGATAGCCACCATCCTGGACTGCCATGCAGC CACCCTGCAGCGGAAGGCTGATCGAGAGGTCTTCTTCATCAACACCCAGAGCATCGTG CAGTTGGTCCAGAGATACCGCAGCGGGATCCGCGGCTATATGAAAACAGTGGTGTTGG ATCTCCTGAGAAGATACTTGCGTGTTGAGAGCAAGGCAAGAGATGCTGATGCCAACAC CAGTGGGATGGTGGGGGGCGTGAGGAGCCTGAGCTTTACCTCTGTGTGGTGTTTTGTC TCCCCCGAATCCCACTACGACAAGTGTGTGATAAACCTCAGGGAGCAGTTCAAGCCAG ACATGTCCCAGGTGCTGGACTGCATCTTCTCCCACGCACAGGTGGCCAAGAAGAACCA GCTGGTGATCATGTTGATCGATGAGCTGTGTGGCCCAGACCCTTCCCTGTCGGACGAG CTGATCTCCATCCTCAACGAGCTCACTCAGCTGAGCAAAAGCGAGCACTGCAAAGTGG

316

| | CCCTCAGAGCCCGGCAGATCCTGATTGCCTCCCACCTCCCCTCCTACGAGCTGCGGCA
TAACCAGGTGGAGTCCATTTTCCTGTCTGCCATTGACATGTACGGCCACCAGTTCTGC
CCCGAGAACCTCAAGAAATTAATACTTTCGGAAACAACCATCTTCGACGTCCTGCCTA
CTTTCTTCTATCACGCAAACAAAGTCGTGTGCATGGCGTCCTTGGAGGTTTACGTGCG
GAGGGGCTACATCGCCTATGAGTTAAACAGCCTGCAGCACCGGCAGCTCCCGGACGGC
ACCTGCGTGGTAGAATTCCAGTTCATGCTGCCGTCCTCCCACCCAAACCGGATGACCG
TGCCCATCAGCATCACCAACCCTGACCTGCTGAGGCACAGCACAGAGCTCTTCATGGA
CAGCGGCTTCTCCCCACTGTGCCAGCGCATGGGAGCCATGGTAGCCTTCAGGAGATTC
GAGGACTTCACCAGAAATTTTGATGAAGTCATCTCTTGCTTCGCCAACGTGCCCAAAG
ACACCCCCCTCTTCAGCGAGGCCCGCACCTCCCTATACTCCGAGGATGACTGCAAGAG
CCTCAGAGAAGAGCCCATCCACATTCTGAATGTGTCCATCCAGTGTGCAGACCACCTG
GAGGATGAGGCACTGGTGCCGATTTTACGGACATTCGTACAGTCCAAGAAAAATATCC
TTGTGGATTATGGACTCCGACGAATCACATTCTTGATTGCCCAAGAGTTTGCAGAAGA
TCGCATTTACCGTCACTTGGAACCTGCCCTGGCCTTCCAGCTGGAACTTAACCGGATG
CGTAACTTCGATCTGACCGCCGTGCCCTGTGCCAACCACAAGATGCACCTTTACCTGG
GTGCTGCCAAGGTGAAGGAAGGTGTGGAAGTGACGGACCATAGGTTCTTCATCCGCGC
CATCATCAGGCACTCTGACCTGATCACAAAGGAAGCCTCCTTCGAATACCTGCAGAAC
GAGGGTGAGCGGCTGCTCCTGGAGGCCATGGACGAGCTGGAGGTGGCGTTCAATAACA
CCAGCGTGCGCACCGACTGCAACCACATCTTCCTCAACTTCGTGCCCACTGTCATCAT
GGACCCCTTCAAGATCGAGGAGTCCGTGCGCTACATGGTTATGCGCTACGGCAGCCGG
CTGTGGAAACTCCGTGTGCTACAGGCTGAGGTCAAGATCAACATCCGCCAGACCACCA
CCGGCAGTGCCGTTCCCATCCGCCTGTTCATCACCAATGAGTCGGGCTACTACCTGGA
CATCAGCCTCTACAAAGAAGTGACTGACTCCAGATCTGGAAATATCATGTTTCACTCC
TTCGGCAACAAGCAAGGGCCCCAGCACGGGATGCTGATCAATACTCCCTACGTCACCA
AGGATCTGCTCCAGCCCAAGCGATTCCAGGCCCAGACCCTGGGAACCACCTACATCTA
TGACTTCCCGGAAATGTTCAGGCAGGCAAGTCCGGCGGCTCAGACGCGGGTACATGTG
CACAATGTGCAGGCTCTCTTTAAACTGTGGGGCTCCCCAGACAAGTATCCCAAAGACA
TCCTGACATACACTGAATTAGTGTTGGACTCTCAGGGCCAGCTGGTGGAGATGAACCG
ACTTCCTGGTGGAAATGAGGTGGGCATGGTGGCCTTCAAAATGAGGTTTAAGCACCAG
GAGTACCCGGAAGGACGGGATGTGATCGTCATCGGCAATGACATCACCTTTCGCATTG
GATCCTTTGGCCCTGGAGAGGACCTTCTGTACCTGCGGGCATCCGAGATGGCCCGGGC
AGAGGGCATTCCCAAAATTTACGTGGCAGCCAACAGTGGCGCCCGTATTGGCATGGCA
GAGGAGATCAAACACATGTTCCACGTGGCTTGGGTGGACCCAGAAGACCCCCACAAAA
AAAAAAAAACAGTGGCTTTCAGTGCAGGGAACTGGATTCGTAGCCTCACTAAAGTATT
TTTTAAGGGATTTAAATACCTGTACCTGACTCCCCAAGACTACACCAGAATCAGCTCC
CTGAACTCCGTCCACTGTAAACACATCGAGGAAGGAGGAGAGTCCAGATACATGATCA
CGGATATCATCGGGAAGGATGATGGCTTGGGCGTGGGAGAATCTGAGGGGCTCAGGCAT
GATTGCTGGGGAGTCCTCTCTGGCTTACGAAGAGATCGTCACCATTAGCTTGGTGACC
TGCCGAGCCATTGGGATTGGGGCCTACTTGGTGAGGCTGGGCCAGCGAGTGATCCAGG
TGGAGAATTCCCACATCATCCTCACAGGAGCAAGTGCTCTCAACAAGGTCCTGGGAAG
AGAGGTCTACACATCCAACAACCAGCTGGGTGGCGTTCAGATCATGCATTACAATGGT
GTCTCCCACATCACCGTGCCAGATGACTTTGAGGGGGTTTATACCATCCTGGAGTGGC
TGTCCTATATGCCAAAGGATAATCACAGCCCTGTCCCTATCATCACACCCACTGACCC
CATTGACAGAGAAATTGAATTCCTCCCATCCAGAGCTCCCTACGACCCCCGGTGGATG
CTTGCAGGAAGGCCTCACCCAACTCTGAAGGGAACGTGGCAGAGCGGATTCTTTGACC
ACGGCAGTTTCAAGGAAATCATGGCACCCTGGGCGCAGACCGTGGTGACAGGACGAGC
AAGGCTTGGGGGGATTCCCGTGGGAGTGATTGCTGTGGAGACACGGACTGTGGAGGTG
GCAGTCCCTGCAGACCCTGCCAACCTGGATTCTGAGGCCAAGATAATTCAGCAGGCAG
GACAGGTGTGGTTCCCAGACTCAGCCTACAAAACCGCCCAGGCCGTCAAGGACTTCAA
CCGGGAGAAGTTGCCCCTGATGATCTTTGCCAACTGGAGGGGGTTCTCCGGTGGCATG
AAAGACATGTATGACCAGGTGCTGAAGTTTGGAGCCTACATCGTGGACGGCCTTAGAC
AATACAAACAGCCCATCCTGATCTATATCCCGCCCTATGCGGAGCTCCGGGGAGGCTC
CTGGGTGGTCATAGATGCCACCATCAACCCGCTGTGCATAGAAATGTATGCAGACAAA
GAGAGCAGGGGTGGTGTTCTGGAACCAGAGGGGACAGTGGAGATTAAGTTCCGAAAGA
AAGATCTGATAAAGTCCATGAGAAGGATCGATCCAGCTTACAAGAAGCTCATGGAACA
GCTAGGGGAACCTGATCTCTCCGACAAGGACCGAAAGGACCTGGAGGGCCGGCTAAAG
GCTCGCGAGGACCTGCTGCTCCCCATCTACCACCAGGTGGCGGTGCAGTTCGCCGACT
TCCATGACACACCCGGCCGGATGCTGGAGAAGGGCGTCATATCTGACATCCTGGAGTG
GAAGACCGCACGCACCTTCCTGTATTGGCGTCTGCGCCGCCTCCTCCTGGAGGACCAG
GTCAAGCAGGAGATCCTGCAGGCCAGCGGGGAGCTGAGTCACGTGCATATCCAGTCCA
TGCTGCGTCGCTGGTTCGTGGAGACGGAGGGGGCTGTCAAGGCCTACTTGTGGGACAA
CAACCAGGTGGTTGTGCAGTGGCTGGAACAGCACTGGCAGGCAGGGGATGGCCCGCGC
TCCACCATCCGTGAGAACATCACGTACCTGAAGCACGACTCTGTCCTCAAGACCATCC
GAGGCCTGGTTGAAGAAACCCCGAGGTGGCCGTGGACTGTGTGATATACCTGAGCCA
GCACATCAGCCCAGCTGAGCGGGCGCAGGTCGTTCACCTGCTGTCTACCATGGACAGC
CCGGCCTCCACC**TGA** |
| | |

	ORF Start: ATG at 1	**ORF Stop: TGA at 7495**	
	SEQ ID NO: 218	**2498 aa**	**MW at 280484.4kD**
NOV76a, **CG59641-01 Protein Sequence**	MVLLLCLSCLIFSCLTFSWLKIWGKMTDSKPITKSKSEANLIPSQEPFPASDNSGETP QRNGEGHTLPKTPSQAEPASHKGPKDAGRRRNSLPPSHQKPPRNPLSSSDAAPSPELQ ANGTGTQGLEATDTNGLSSSARPQGQQAGSPSKEDKKQANIKRQLMTNFILGSFDDYS SDEDSVAGSSRESTRKGSRASLGALSLEAYLTTRPSMSGLHLVKRGREHKKLDLHRDF TVASPAEFVTRFGGDRVIEKVLIANNGIAAVKCMRSIRRWAYEMFRNERAIRFVVMVT PEDLKANAEYIKMADHYVPVPGGPNNNNYANVELIVDIAKRIPVQAVWAGWGHASENP		

```
KLPELLCKNGVAFLGPPSEAMWALGDKIASTVVAQTLQVPTLPWSGSGLTVEWTEDDL
QQGKRISVPEDVYDKGCVKDVDEGLEAAERIGFPLMIKASEGGGGKGIRKAESAEDFP
ILFRQVQSEIPGSPIFLMKLAQHARHLEVQILADQYGNAVSLFGRDCSIQRRHQKIVE
EAPATIAPLAIFEFMEQCAIRLAKTVGYVSAGTVEYLYSQDGSFHFLELNPRLQVEHP
CTEMIADVNLPAAQLQIAMGVPLHRLKDIRLLYGESPWGVTPISFETPSNPPLARGHV
IAARITSENPDEGFKPSSGTVQELNFRSSKNVWGYFSVAATGGLHEFADSQFGHCFSW
GENREEAISNMVVALKELSIRGDFRTTVEYLINLLETESFQNNDIDTGWLDYLIAEKV
QAEKPDIMLGVVCGALNVADAMFRTCMTDFLHSLERGQVLPADSLLNLVDVELIYGGV
KYILKVARQSLTMFVLIMNGCHIEIDAHRLNDGGLLLSYNGNSYTTYMKEEVDSYRIT
IGNKTCVFEKENDPTVLRSPSAGKLTQYTVEDGGHVEAGSSYAEMEVMKMIMTLNVQE
RGRVKYIKRPGAVLEAGCVVARLELDDPSKVHPAEPFTGELPAQQTLPILGEKLHQVF
HSVLENLTNVMSGFCLPEPVFSIKLKEWVQKLMMTLRHPSLPLLELQEIMTSVAGRIP
APVEKSVRRVMAQYASNITSVLCQFPSQQIATILDCHAATLQRKADREVFFINTQSIV
QLVQRYRSGIRGYMKTVVLDLLRRYLRVESKARDADANTSGMVGGVRSLSFTSVWCFV
SPESHYDKCVINLREQFKPDMSQVLDCIFSHAQVAKKNQLVIMLIDELCGPDPSLSDE
LISILNELTQLSKSEHCKVALRARQILIASHLPSYELRHNQVESIFLSAIDMYGHQFC
PENLKKLILSETTIFDVLPTFFYHANKVVCMASLEVYVRRGYIAYELNSLQHRQLPDG
TCVVEFQFMLPSSHPNRMTVPISITNPDLLRHSTELFMDSGFSPLCQRMGAMVAFRRF
EDFTRNFDEVISCFANVPKDTPLFSEARTSLYSEDDCKSLREEPIHILNVSIQCADHL
EDEALVPILRTFVQSKKNILVDYGLRRITFLIAQEFAEDRIYRHLEPALAFQLELNRM
RNFDLTAVPCANHKMHLYLGAAKVKEGVEVTDHRFFIRAIIRHSDLITKEASFEYLQN
EGERLLLEAMDELEVAFNNTSVRTDCNHIFLNFVPTVIMDPFKIEESVRYMVMRYGSR
LWKLRVLQAEVKINIRQTTTGSAVPIRLFITNESGYYLDISLYKEVTDSRSGNIMFHS
FGNKQGPQHGMLINTPYVTKDLLQAKRFQAQTLGTTYIYDFPEMFRQASPAAQTRVHV
HNVQALFKLWGSPDKYPKDILTYTELVLDSQGQLVEMNRLPGGNEVGMVAFKMRFKTQ
EYPEGRDVIVIGNDITFRIGSFGPGEDLLYLRASEMARAEGIPKIYVAANSGARIGMA
EEIKHMFHVAWVDPEDPHKKKKTVAFSAGNWIRSLTKVFFKGFKYLYLTPQDYTRISS
LNSVHCKHIEEGGESRYMITDIIGKDDGLGVENLRGSGMIAGESSLAYEEIVTISLVT
CRAIGIGAYLVRLGQRVIQVENSHIILTGASALNKVLGREVYTSNNQLGGVQIMHYNG
VSHITVPDDFEGVYTILEWLSYMPKDNHSPVPIITPTDPIDREIEFLPSRAPYDPRWM
LAGRPHPTLKGTWQSGFFDHGSFKEIMAPWAQTVVTGRARLGGIPVGVIAVETRTVEV
AVPADPANLDSEAKIIQQAGQVWFPDSAYKTAQAVKDFNREKLPLMIFANWRGFSGGM
KDMYDQVLKFGAYIVDGLRQYKQPILIYIPPYAELRGGSWVVIDATINPLCIEMYADK
ESRGGVLEPEGTVEIKFRKKDLIKSMRRIDPAYKKLMEQLGEPDLSDKDRKDLEGRLK
AREDLLLPIYHQVAVQFADFHDTPGRMLEKGVISDILEWKTARTFLYWRLRRLLLEDQ
VKQEILQASGELSHVHIQSMLRRWFVETEGAVKAYLWDNNQVVVQWLEQHWQAGDGPR
STIRENITYLKHDSVLKTIRGLVEENPEVAVDCVIYLSQHISPAERAQVVHLLSTMDS
PAST
```

Further analysis of the NOV76a protein yielded the following properties shown in Table 76B.

Table 76B. Protein Sequence Properties NOV76a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 25 and 26

A search of the NOV76a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 76C.

318

Table 76C. Geneseq Results for NOV76a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV76a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU32848	Novel human secreted protein #3339 - Homo sapiens, 2486 aa. [WO200179449-A2, 25-OCT-2001]	26..2498 1..2486	2316/2555 (90%) 2339/2555 (90%)	0.0
AAR05707	Acetyl-CoA-carboxylase - Gallus sp, 2324 aa. [JP02057179-A, 26-FEB-1990]	163..2498 17..2324	1728/2375 (72%) 2003/2375 (83%)	0.0
AAB86033	Bovine acetyl-coenzyme A carboxylase-alpha protein fragment - Bos taurus, 2288 aa. [DE19946173-A1, 05-APR-2001]	204..2497 14..2288	1719/2342 (73%) 1969/2342 (83%)	0.0
AAR98811	Erysiphe graminis acetyl coenzyme A carboxylase - Erysiphe graminis f.sp.hordei, 2273 aa. [FR2727129-A1, 24-MAY-1996]	235..2490 42..2271	1045/2326 (44%) 1432/2326 (60%)	0.0
AAY24150	Candida albicans acetyl CoA carboxylase - Candida albicans, 2270 aa. [WO9932635-A1, 01-JUL-1999]	239..2489 88..2269	1015/2300 (44%) 1396/2300 (60%)	0.0

In a BLAST search of public sequence databases, the NOV76a protein was found to have homology to the proteins shown in the BLASTP data in Table 76D.

Table 76D. Public BLASTP Results for NOV76a				
Protein Accession Number	Protein/Organism/Length	NOV76a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O00763	Acetyl-CoA carboxylase 2 (EC 6.4.1.2) (ACC-beta) [Includes: Biotin carboxylase (EC 6.3.4.14)] - Homo sapiens (Human), 2483 aa.	1..2498 1..2483	2349/2528 (92%) 2384/2528 (93%)	0.0
O70151	ACETYL-COA CARBOXYLASE - Rattus norvegicus (Rat), 2456 aa.	1..2497 1..2455	2068/2524 (81%) 2224/2524 (87%)	0.0

319

CAA48770	ACETYL-COA CARBOXYLASE (EC 6.4.1.2) - Homo sapiens (Human), 2339 aa.	163..2498 17..2339	1921/2390 (80%) 2086/2390 (86%)	0.0
P11029	Acetyl-CoA carboxylase (EC 6.4.1.2) (ACC) [Includes: Biotin carboxylase (EC 6.3.4.14)] - Gallus gallus (Chicken), 2324 aa.	163..2498 17..2324	1732/2375 (72%) 2004/2375 (83%)	0.0
P11497	Acetyl-CoA carboxylase 1 (EC 6.4.1.2) (ACC-alpha) [Includes: Biotin carboxylase (EC 6.3.4.14)] - Rattus norvegicus (Rat), 2345 aa.	163..2497 17..2345	1736/2396 (72%) 1993/2396 (82%)	0.0

PFam analysis predicts that the NOV76a protein contains the domains shown in the Table 76E.

Table 76E. Domain Analysis of NOV76a			
Pfam Domain	NOV76a Match Region	Identities/ Similarities for the Matched Region	Expect Value
CPSase_L_chain: domain 1 of 1	249..372	49/132 (37%) 117/132 (89%)	2.2e-57
CPSase_L_D2: domain 1 of 1	374..619	102/253 (40%) 218/253 (86%)	6.6e-118
Biotin_carb_C: domain 1 of 1	640..747	40/118 (34%) 100/118 (85%)	1.9e-53
biotin_lipoyl: domain 1 of 1	885..951	22/75 (29%) 56/75 (75%)	6.5e-17
Carboxyl_trans: domain 1 of 2	1783..1878	31/100 (31%) 88/100 (88%)	7.4e-34
GTP_cyclohydroI: domain 1 of 1	2287..2304	6/18 (33%) 13/18 (72%)	6.6
Carboxyl_trans: domain 2 of 2	1897..2374	191/504 (38%) 447/504 (89%)	4.1e-258

Example 77.

The NOV77 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 77A.

Table 77A. NOV77 Sequence Analysis		
	SEQ ID NO: 219	1624 bp
NOV77a, CG59630-01 DNA Sequence	CGCGCGCCGGGG**ATG**GAGCCGCAGCCCGGCGGCGCCCGGAGCTGCCGGCGCGGGGCCC CCGGCGGCGCCTGCGAGCTGGGCCCGGCGGCCGAGGCGGCGCCCATGAGCCTCGCCAT CCACAGCACCACGGGCACCCGCTACGACCTGGCCGTGCCGCCCGACGAGACGGTGGAG GGGCTGCGCAAGCGGTTGTCCCAGCGCCTCAAAGTGCCCAAGGAGCGCCTGGCTCTTC TCCACAAAGACACGCGGCTCAGTTCGGGGAAGCTGCAGGAGTTCGGCGTGGGTGATGG CAGCAAGCTGACCTTGGTACCCACCGTGGAAGCGGGCCTCATGTCTCAGGCCTCAAGG CCGGAACAGTCCGTGATGCAAGCTCTCGAGAGTCTCACGGAGACGCAGCCCCCAGCGG CGCCCGGGCCGGGCCGGGCTGGCGGAGGAGGCTTCCGGAAATACAGATTCATTTTATT TAAGCGTCCGTGGCACCGACAGGGACCCCAGAGCCCAGAGAGGGGCGGCGAGAGGCCC CAGGTCAGTGACTTCCTGTCGGGCCGTTCGCCACTGACACTGGCCTTGCGTGTGGGCG ACCACATGATGTTCGTGCAGCTGCAGCTCGCGGCCCAGCACGCTCCACTGCAACACCG CCATGTGCTGGCCGCTGCGGCCGCCGCCGCTGCTGCGCGGGGGGACCCCAGCATAGCC TCCCCCGTGTCCTCGCCCTGCCGGCCGGTGTCCAGTGCCGCCCGAGTCCCCCCGGTGC CCACCAGCCCGTCCCCTGCATCTCCCTCGCCCATCACAGCCGGCTCCTTCCGGTCCCA CGCAGCCTCCACCACCTGCCCGGAGCAGATGGACTGCTCCCCCACGGCCAGCAGCAGT GCCAGTCCTGGTGCCAGCACCACGTCTACCCCAGGGGCCAGCCCTGCCCCCCCGCTCCC GAAAACCCGGCGCCGTCATCGAGAGCTTTGTGAATCACGCCCCGGGGGTCTTCTCAGG GACCTTCTCTGGCACGCTACACCCCAACTGCCAAGACAGCAGCGGGCGGCCGCGGCGT GACATCGGCACCATCCTGCAGATCCTGAACGACCTCCTGAGCGCCACCCGGCACTACC AGGGCATGCCCCCTTCGCTGGCCCAGCTCCGCTGCCACGCCCAGTGCTCCCCGGCCTC ACCGGCCCCCGACCTGGCCCCCAGAACTACCTCCTGCGAGAAGCTCACGGCTGCCCCC TCAGCCTCCCTGCTGCAGGGCCAGAGCCAGATCCGCATGTGCAAGCCCCCGGGTGACC GGCTTCGGCAGACAGAAACCGCGCCACGCGCTGCAAGGTGGAACGGCTGCAGCTGCT TCTGCAGCAGAAACGGCTCCGTAGAAAGGCCCGGCGGGACGCGCGGGGTCCGTACCAC TGGTCACCCAGCCGCAAGGCCGGCCGCAGCGACAGCAGTAGCAGCGGGGGCGGCGGCA GCCCCAGCGAGGCCTCCGGCTTGGGCCTCGACTTCGAGGACTCCGTGTGGAAGCCAGA AGTCAACCCTGACATCAAGTCAGAGTTCGTGGTGGCT**TAG**GATCTTCGGATCGGCCAC CCTCGCCCCTCGCACCCCAGCCCAGGGCGGCGGGGACTCCGAGAGCCCCGGAGAGAAC
	ORF Start: ATG at 13	ORF Stop: TAG at 1546
	SEQ ID NO: 220	511 aa	MW at 53949.3kD
NOV77a, CG59630-01 Protein Sequence	MEPQPGGARSCRRGAPGGACELGPAAEAAPMSLAIHSTTGTRYDLAVPPDETVEGLRK RLSQRLKVPKERLALLHKDTRLSSGKLQEFGVGDGSKLTLVPTVEAGLMSQASRPEQS VMQALESLTETQPPAAPGPGRAGGGGFRKYRFILFKRPWHRQGPQSPERGGERPQVSD FLSGRSPLTLALRVGDHMMFVQLQLAAQHAPLQHRHVLAAAAAAAAARGDPSIASPVS SPCRPVSSAARVPPVPTSPSPASPSPITAGSFRSHAASTTCPEQMDCSPTASSSASPG ASTTSTPGASPAPRSRKPGAVIESFVNHAPGVFSGTFSGTLHPNCQDSSGRPRRDIGT ILQILNDLLSATRHYQGMPPSLAQLRCHAQCSPASPAPDLAPRTTSCEKLTAAPSASL LQGQSQIRMCKPPGDRLRQTENRATRCKVERLQLLLQQKRLRRKARRDARGPYHWSPS RKAGRSDSSSSGGGGSPSEASGLGLDFEDSVWKPEVNPDIKSEFVVA

Further analysis of the NOV77a protein yielded the following properties shown in Table 77B.

Table 77B. Protein Sequence Properties NOV77a	
PSort analysis:	0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1526 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV77a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 77C.

321

Table 77C. Geneseq Results for NOV77a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV77a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB56832	Human prostate cancer antigen protein sequence SEQ ID NO:1410 - Homo sapiens, 236 aa. [WO200055174-A1, 21-SEP-2000]	267..493 1..227	189/227 (83%) 195/227 (85%)	e-104

In a BLAST search of public sequence databases, the NOV77a protein was found to have homology to the proteins shown in the BLASTP data in Table 77D.

Table 77D. Public BLASTP Results for NOV77a				
Protein Accession Number	Protein/Organism/Length	NOV77a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9JJJ6	MIDNOLIN - Mus musculus (Mouse), 508 aa.	1..511 1..508	475/514 (92%) 486/514 (94%)	0.0
Q96BW8	SIMILAR TO MIDNOLIN - Homo sapiens (Human), 177 aa (fragment).	338..511 4..177	174/174 (100%) 174/174 (100%)	2e-97
Q9W2S4	CG9732 PROTEIN - Drosophila melanogaster (Fruit fly), 989 aa.	213..363 524..677	58/155 (37%) 80/155 (51%)	6e-18
AAL40834	BPLF1 - Human herpesvirus 4 (Epstein-Barr virus), 3179 aa.	200..406 320..530	64/223 (28%) 95/223 (41%)	2e-07
Q9BKV7	PPG3 - Leishmania major, 1325 aa.	213..328 984..1104	37/121 (30%) 66/121 (53%)	2e-06

PFam analysis predicts that the NOV77a protein contains the domains shown in the

5   Table 77E.

322

Table 77E. Domain Analysis of NOV77a			
Pfam Domain	**NOV77a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ubiquitin: domain 1 of 1	31..99	19/79 (24%) 46/79 (58%)	0.00033
PI3_PI4_kinase: domain 1 of 1	411..427	7/18 (39%) 14/18 (78%)	1.5

Example 78.

The NOV78 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 78A.

Table 78A. NOV78 Sequence Analysis			
	SEQ ID NO: 221	1034 bp	
NOV78a, CG59561-01 DNA Sequence	CCACCGCCACAGCTGCCAGC**ATG**TCTGGCCCAGACATCAAGACGCCGACCGCCATCCA GATCTGCCGGATTATGCGGGACGCTAATGTGGCCCGCAATGTCTACGGCGGGACCATC CTGAAGATGATCAAAGAGGCGGGCGCCATCATCAGCACCCGGCATTGCAATCCGCAGA ACGGGGATCGCTGTGTGGCCGCTCTGGCTCGGGTCGAGTGCACCCACTTCCTGTGGCC CATGTGCATCGGTGAGGTGGCCCACGTCAGCGCGGAGATCACCTACACCTCCAAGCAC TCTGTGGAGGTGCAGGTCAACATGATGTCCGAAAACATCCTCACAGGTGCCAAAAAGC TGACCAATAAGGCCACCCTCTGGTATGCGCCCCTGTCGCTGACGAACGTGGACAAGGT CCTCGAAGAGCCTCCTGTTGTGTATTCCGGCAGGAGCAGGAGGAGGAGGGCCAGAAG CGGTACAAAACCCAGAAGCTGGAGCGCATGGAGACCAACTGGAGGAACGGGGACATCG TCCAGCCAGTCCTCAACCCAGAGCCGAACACTGTCAGCTACAGCCAGTCCAGCTTGAT CCACCTGGTGGGGCCTTCAGACTGTACCCTGCACAGCTTCGTGCATGAAGGGGGTGACC ATGAAGGTCATGGACGAGGTCGCCGGGATCTTGGCTGCACGCCACTGCAAGACCAACC TCGTCACAGCCTCCATGGAGGCCATTAATTTTGACAACAAGATCAGAAAAGGCTGCAT CAAGACCATCTCCGGACGCATGACCTTCACGAGCAATAAGTCCGTAGAGATCGAGGTC TTGGTGGATGCCGACTGTGTTGTGGACAGCTCTCAGAAGCGCTACAGGGCCGCCAGTG TCTTCACCTATGTGTCGCTGAGCCAGGAAGGCAGGTCGCTGCCCATGCCCCAGCTCGT GCCGGAGACCCAGGACGAGAAGGGCTTTGAGGCCTGGCTCGGTGGCTCACGCCTA**TAA** TCCCAGCACTTTAGGATGCTGAGGCAGGCGGATCACTTGACGTCAGGA		
	ORF Start: ATG at 21	ORF Stop: TAA at 984	
	SEQ ID NO: 222	321 aa	MW at 35738.7kD
NOV78a, CG59561-01 Protein Sequence	MSGPDIKTPTAIQICRIMRDANVARNVYGGTILKMIKEAGAIISTRHCNPQNGDRCVA ALARVECTHFLWPMCIGEVAHVSAEITYTSKHSVEVQVNMMSENILTGAKKLTNKATL WYAPLSLTNVDKVLEEPPVVYFRQEQEEEGQKRYKTQKLERMETNWRNGDIVQPVLNP EPNTVSYSQSSLIHLVGPSDCTLHSFVHEGVTMKVMDEVAGILAARHCKTNLVTASME AINFDNKIRKGCIKTISGRMTFTSNKSVEIEVLVDADCVVDSSQKRYRAASVFTYVSL SQEGRSLPMPQLVPETQDEKGFEAWLGGSRL		

Further analysis of the NOV78a protein yielded the following properties shown in Table 78B.

Table 78B. Protein Sequence Properties NOV78a	
PSort analysis:	0.8000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV78a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 78C.

Table 78C. Geneseq Results for NOV78a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV78a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW74896	Human secreted protein encoded by gene 169 clone HPTTU11 - Homo sapiens, 339 aa. [WO9839448-A2, 11-SEP-1998]	1..310 1..313	273/313 (87%) 292/313 (93%)	e-154
AAY71115	Human Hydrolase protein-13 (HYDRL-13) - Homo sapiens, 375 aa. [WO200028045-A2, 18-MAY-2000]	1..310 33..316	247/313 (78%) 266/313 (84%)	e-133
AAY35275	Chlamydia pneumoniae transmembrane protein sequence - Chlamydia pneumoniae, 155 aa. [WO9927105-A2, 03-JUN-1999]	187..310 16..138	35/124 (28%) 72/124 (57%)	1e-09
AAG92590	C glutamicum protein fragment SEQ ID NO: 6344 - Corynebacterium glutamicum, 339 aa. [EP1108790-A2, 20-JUN-2001]	24..309 35..307	69/296 (23%) 112/296 (37%)	7e-08
AAB76624	Corynebacterium glutamicum MCT protein SEQ ID NO:230 - Corynebacterium glutamicum, 339 aa. [WO200100805-A2, 04-JAN-2001]	24..309 35..307	69/296 (23%) 112/296 (37%)	7e-08

In a BLAST search of public sequence databases, the NOV78a protein was found to have homology to the proteins shown in the BLASTP data in Table 78D.

Table 78D. Public BLASTP Results for NOV78a				
Protein Accession Number	Protein/Organism/Length	NOV78a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O00154	Cytosolic acyl coenzyme A thioester hydrolase (EC 3.1.2.2) (Long chain acyl-CoA thioester hydrolase) (CTE-II) (Brain acyl-CoA hydrolase) (BACH) - Homo sapiens (Human), 338 aa.	1..310 1..313	274/313 (87%) 293/313 (93%)	e-154
Q91V12	ACYL-COA HYDROLASE (HYPOTHETICAL 37.6 KDA PROTEIN) - Mus musculus (Mouse), 338 aa.	1..310 1..313	265/313 (84%) 287/313 (91%)	e-150
Q64559	Cytosolic acyl coenzyme A thioester hydrolase (EC 3.1.2.2) (Long chain acyl-CoA thioester hydrolase) (CTE-II) (Brain acyl-CoA hydrolase) (BACH) (ACT) (LACH1) (ACH1) - Rattus norvegicus (Rat), 338 aa.	1..310 1..313	263/313 (84%) 286/313 (91%)	e-149
JC5416	palmitoyl-CoA hydrolase (EC 3.1.2.2), hepatic - rat, 343 aa.	12..310 17..318	251/302 (83%) 276/302 (91%)	e-142
Q9Y541	DJ202O8.3.1 (HBACH (BRAIN ACYL-COA HYDROLASE (ACYL COENZYME A THIOESTER HYDROLASE, EC 3.1.2.2)) (ISOFORM 1)) - Homo sapiens (Human), 237 aa (fragment).	1..202 33..236	181/204 (88%) 190/204 (92%)	e-100

PFam analysis predicts that the NOV78a protein contains the domains shown in the Table 78E.

Table 78E. Domain Analysis of NOV78a			
Pfam Domain	NOV78a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Acyl-CoA_hydro: domain 1 of 1	165..305	46/147 (31%) 131/147 (89%)	1.1e-47

Example 79.

The NOV79 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 79A.

325

SEQ ID NO: 223	4203 bp
NOV79a, CG59452-01 DNA Sequence	AATGTGATGGGATCACTAGC**ATG**TCTGCGGGAGAGCGGCCCTGGGACGAGATTGAGAAA TCTGCCAGTAATGGGGGATGGACTAGAAACTTCCCAAATGTCTACAACACAGGCCCAG GCCCAACCCCAGCCAGCCAACGCAGCCAGCACCAACCACCAACCCCCGCCCCCAGAGACCTCCA ACCCTAACAAGCCCAAGAGGCAGACCAACCAACTGCAATACCTGCTCAGAGTGGTGCT CAAGACACTATGGAAACACCAGTTTGCATGGCCTTTCCAGCAGCCTGTGGATGCCGTC AAGCTGAACCTCCCTGATTACTATAAGATCATTAAAACGCCTATGGATATGGGAACAA TAAAGAAGCGCTTGGAAAACAACTATTACTGGAATGCTCAGGAATGCTGTATCCAGGACTT CAACACTATGTTTACAAATTGTTACATCTACAACAAGCCTGGAGATGACATAGTCTTA ATGGCAGAAGCTCTGGAAAAGCTCTTCTTGCAAAAAATAAATGAGCTACCCACAGAAG AAACCGAGATCATGATAGTCCAGGCAAAAGGAAGAGGACGTGGGAGGAAAGAAACAGG TACAGCAAAACCTGGCGTTTCCACGGTACCAAACACAACTCAAGCATCGACTCCTCCG CAGACCCAGACCCCTCAGCCGAATCCTCCTCCTGTGCAGGCCACGCCTCACCCCTTCC CTGCCGTCACCCCGGACCTCATCGTCCAGACCCCTGTCATGACAGTGGTGCCTCCCCA GCCACTGCAGACGCCCCCGCCAGTGCCCCCCCAGCCACAACCCCCACCCGCTCCAGCT CCCCAGCCCGTACAGAGCCACCCACCCATCATCGCGGCCACCCCACAGCCTGTGAAGA CAAAGAAGGGAGTGAAGAGGAAAGCAGACACCACCACCCCCACCACCATTGACCCCAT TCACGAGCCACCCTCGCTGCCCCCGGAGCCCAAGACCACCAAGCTGGGCCAGCGGCGG GAGAGCAGCCGGCCTGTGAAACCTCCAAAGAAGGACGTGCCCGACTCTCAGCAGCACC CAGCACCAGAGAAGAGCAGCAAGGTCTCGGAGCAGCTCAAGTGCTGCAGCGGCATCCT CAAGGAGATGTTTGCCAAGAAGCACGCCGCCTACGCCTGGCCCTTCTACAAGCCTGTG GACGTGGAGGCACTGGGCCTACACGACTACTGTGACATCATCAAGCACCCCATGGACA TGAGCACAATCAAGTCTAAACTGGAGGCCCGTGAGTACCGTGATGCTCAGGAGTTTGG TGCTGACGTCCGATTGATGTTCTCCAACTGCTATAAGTACAACCCTCCTGACCATGAG GTGGTGGCCATGGCCCGCAAGCTCCAGGATGTGTTCGAAATGCGCTTTGCCAAGATGC CGGACGAGCCTGAGGAGCCAGTGGTGGCCGTGTCCTCCCCGGCAGTGCCCCCTCCCAC CAAGGTTGTGGCCCCGCCCTCATCCAGCGACAGCAGCAGCGATAGCTCCTCGGACAGT GACAGTTCGACTGATGACTCTGAGGAGGAGCGAGCCCAGCGGCTGGCTGAGCTCCAGG AGCAGCTCAAAGCCGTGCACGAGCAGCTTGCAGCCCTCTCTCAGCCCCAGCAGAACAA ACCAAAGAAAAGGAGAAAGACAAGAAGGAAAAGAAAAAAGAAAAGCACAAAAGGAAA GAGGAAGTGGAAGAGAATAAAAAAAAGCAAAGCCAAGGAACCTCCTCCTAAAAAGACGA AGAAAAATAATAGCAGCAACAGCAATGTGAGCAAGAAGGAGCCAGCGCCCATGAAGAG CAAGCCCCCTCCCACGTATGAGTCGGAGGAAGAGGACAAGTGCAAGCCTATGTCCTAT GAGGAGAAGCGGCAGCTCAGCTTGGACATCAACAAGCTCCCCGGCGAGAAGCTGGGCC GCGTGGTGCACATCATCCAGTCACGGGAGCCCTCCCTGAAGAATTCCAACCCCGACGA GATTGAAATCGACTTTGAGACCCTGAAGCCGTCCACACTGCGTGAGCTGGAGCGCTAT GTCACCTCCTGTTTGCGGAAGAAAAGGAAACCTCAAGCTGAGAAAGTTGATGTGATTG CCGGCTCCTCCAAGATGAAGGGCTTCTCGTCCTCAGAGTCGGAGAGCTCCAGTGAGTC CAGCTCCTCTGACAGCGAAGACTCCGAAACAGAGATGGCTCCGAAGTCAAAAAAGAAG GGGCACCCGGGAGGGAGCAGAAGCAGCACCATCATCACCACCATCAGCAGATGCAGC AGGCCCCGGCTCCTGTGCCCCAGCAGCCGCCCCCGCCTCCCCAGCAGCCCCCACCGCC TCCACCTCCGCAGCAGCAACAGCAGCCGCCACCCCCGCCTCCCCCACCCTCCATGCCG CAGCAGGCAGCCCCGGCGATGAAGTCCTCGCCCCCACCCTTCATTGCCACCCAGGTGC CCGTCCTGGAGCCCCAGCTCCCAGGCAGCGTCTTTGACCCCATCGGCCACTTCACCCA GCCCATCCTGCACCTGCCGCAGCCTGAGCTGCCCCCTCACCTGCCCCAGCCGCCTGAG CACAGCACTCCACCCCATCTCAACCAGCACGCAGTGGTCTCTCCTCCAGCTTTGCACA ACGCACTACCCCAGCAGCCATCACGGCCCAGCAACCGAGCCGCTGCCCTGCCTCCCAA GCCCGCCCGGCCCCCAGCCGTGTCACCAGCCTTGACCCAAACACCCCTGCTCCCACAG CCCCCCATGGCCCAACCCCCCCAAGTGCTGCTGGAGGATGAAGAGCCACCTGCCCCAC CCCTCACCTCCATGCAGATGCAGCTGTACCTGCAGCAGCTGCAGAAGGTGCAGCCCCC TACGCCGCTACTCCCTTCCGTGAAGGTGCAGTCCCAGCCCCCACCCCCCCTGCCGCCC CCACCCCACCCCTCTGTGCAGCAGCAGCTGCAGCAGCAGCCGCCACCACCCCCACCAC CCCAGCCCCAGCCTCCACCCCAGCAGCAGCATCAGCCCCCTCCACGGCCCGTGCACTT GCAGCCCATGCAGTTTTCCACCCACATCCAACAGCCCCCGCCACCCCAGGGCCAGCAG CCCCCCCATCCGCCCCCAGGCCAGCAGCCACCCCCGCCCGCAGCCTGCCAAGCCTCAGC AAGTCATCCAGCACCACCATTCACCCCGGCACCACAAGTCGGACCCCTACTCAACCGG TCACCTCCGCGAAGCCCCCTCCCCGCTTATGATACATTCCCCCCAGATGTCACAGTTC CAGAGCCTGACCCACCAGTCTCCACCCCAGCAAAACGTCCAGCCTAAGAAACAGGTAA CTGGCAGGGCTGGGCCAAGTCCTGTGGGCCAGGGCCGGGGGTGCCTGCCCACCTCACC GGCCGCTGTGCCTGTGCCATCCCAGGAGCTGCGTGCTGCCTCCGTGGTCCAGCCCCAG CCCCTCGTGGTGGTGAAGGAGGAGAAGATCCACTCACCCATCATCCGCAGCGAGCCCT TCAGCCCCTCGCTGCGGCCGGAGCCCCCCAAGCACCCGGAGAGCATCAAGGCCCCCGT TTATGTTCCAGGGCCGGAAATGAAGCCTGTGGATGTCGGGAGGCCTGTGATCCGGCCC CCAGAGCAGAACGCACCGCCACCAGGGGCCCCTGACAAGGACAAACAGAAACAGGAGC CGAAGACTCCAGTTGCGCCCAAAAAGGACCTGAAAATCAAGAACATGGGCTCCTGGGC CAGCCTAGTGCAGAAGCATCCGACCACCCCCTCCTCCACAGCCAAGTCATCCAGCGAC AGCTTCGAGCAGTTCCGCCGCGCCGCTCGGGAGAAAGAGGAGCGTGAGAAGGCCCTGA AGGCTCAGGCCGAGCACCTGAGCAAGGAGAAGGAGCGGCTGCGGCAGGAGCGCATGAG GAGCCGAGAGGACGAGGATGCGCTGGAGCAGGGCCCGGCGGGCCCATGAGGAGGCACGT CGGCGCCAGGAGCAGCAGCAGCAGCAGCGCCAGGAGCAACAGCAGCAGCAGCAACAGC AAGCAGCTGCGGTGGCTGCCGCCGCCACCCCACAGGCCCAGAGCTCCCAGCCCCAGTC CATGCTGGACCAGCAGAGGGAGTTGGCCCGGAAGCGGGAGCAGGAGCGAAGACGCCGG GAAGCCATGGCAGCTACCATTGACATGAATTTCCAGAGTGATCTATTGTCAATATTTG AAGAAAATCTTTTC**TGA**GCGCACCTAG

	ORF Start: ATG at 21	ORF Stop: TGA at 4191	
	SEQ ID NO: 224	1390 aa	MW at 154728.4kD
NOV79a, CG59452-01 Protein Sequence	MSAESGPGTRLRNLPVMGDGLETSQMSTTQAQAQPQPANAASTNPPPPETSNPNKPKR QTNQLQYLLRVVLKTLWKHQFAWPFQQPVDAVKLNLPDYYKIIKTPMDMGTIKKRLEN NYYWNAQECIQDFNTMFTNCYIYNKPGDDIVLMAEALEKLFLQKINELPTEETEIMIV QAKGRGRGRKETGTAKPGVSTVPNTTQASTPPQTQTPQPNPPPVQATPHPFPAVTPDL IVQTPVMTVVPPQPLQTPPPVPPQPQPPPAPAPQPVQSHPPIIAATPQPVKTKKGVKR KADTTTPTTIDPIHEPPSLPPEPKTTKLGQRRESSRPVKPPKKDVPDSQQHPAPEKSS KVSEQLKCCSGILKEMFAKKHAAYAWPFYKPVDVEALGLHDYCDIIKHPMDMSTIKSK LEAREYRDAQEFGADVRLMFSNCYKYNPPDHEVVAMARKLQDVFEMRFAKMPDEPEEP VVAVSSPAVPPPTKVVAPPSSSDSSSDSSSDSDSSTDDSEEERAQRLAELQEQLKAVH EQLAALSQPQQNKPKKKEKDKKEKKKEKHKRKEEVEENKKSKAKEPPPKKTKKNNSSN SNVSKKEPAPMKSKPPPTYESEEEDKCKPMSYEEKRQLSLDINKLPGEKLGRVVHIIQ SREPSLKNSNPDEIEIDFETLKPSTLRELERYVTSCLRKKRKPQAEKVDVIAGSSKMK GFSSSESESSSESSSSDSEDSETEMAPKSKKKGHPGREQKQHHHHHHQQMQQAPAPVP QQPPPPPQQPPPPPPPQQQQQPPPPPPPPSMPQQAAPAMKSSPPPFIATQVPVLEPQL PGSVFDPIGHFTQPILHLPQPELPPHLPQPPEHSTPPHLNQHAVVSPPALHNALPQQP SRPSNRAAALPPKPARPPAVSPALTQTPLLPQPPMAQPPQVLLEDEEPPAPPLTSMQM QLYLQQLQKVQPPTPLLPSVKVQSQPPPPLPPPHPSVQQQLQQQPPPPPPPQPQPPP QQQHQPPPRPVHLQPMQFSTHIQQPPPPQGQQPPHPPPGQQPPPPQPAKPQQVIQHHH SPRHHKSDPYSTGHLREAPSPLMIHSPQMSQFQSLTHQSPPQQNVQPKKQVTGRAGPS PVGQGRGCLPTSPAAVPVPSQELRAASVVQPQPLVVVKEEKIHSPIIRSEPFSPSLRP EPPKHPESIKAPVYVPGPEMKPVDVGRPVIRPPEQNAPPPGAPDKDKQKQEPKTPVAP KKDLKIKNMGSWASLVQKHPTTPSSTAKSSSDSFEQFRRAAREKEEREKALKAQAEHA EKEKERLRQERMRSREDEDALEQARRAHEEARRRQEQQQQQRQEQQQQQQQQAAAVAA AATPQAQSSQPQSMLDQQRELARKREQERRRREAMAATIDMNFQSDLLSIFEENLF

Further analysis of the NOV79a protein yielded the following properties shown in Table 79B.

Table 79B. Protein Sequence Properties NOV79a	
PSort analysis:	0.9800 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV79a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 79C.

Table 79C. Geneseq Results for NOV79a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV79a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY57898	Human transmembrane protein HTMPN-22 - Homo sapiens, 688 aa. [WO9961471-A2, 02-DEC-1999]	1..667 1..667	667/667 (100%) 667/667 (100%)	0.0

AAY07027	Breast cancer associated antigen precursor sequence - Homo sapiens, 754 aa. [WO9904265-A2, 28-JAN-1999]	44..724 13..708	407/732 (55%) 487/732 (65%)	0.0
AAY07114	WO9904265 Seq ID No: 685 - Homo sapiens, 947 aa. [WO9904265-A2, 28-JAN-1999]	35..738 4..686	357/761 (46%) 444/761 (57%)	e-170
AAW81168	Transcriptional regulatory factor RING3 - Homo sapiens, 947 aa. [WO9848015-A1, 29-OCT-1998]	35..738 4..686	357/761 (46%) 444/761 (57%)	e-170
AAU16206	Human novel secreted protein, Seq ID 1159 - Homo sapiens, 235 aa. [WO200155322-A2, 02-AUG-2001]	51..255 1..203	118/206 (57%) 137/206 (66%)	2e-59

In a BLAST search of public sequence databases, the NOV79a protein was found to have homology to the proteins shown in the BLASTP data in Table 79D.

Table 79D. Public BLASTP Results for NOV79a				
Protein Accession Number	Protein/Organism/Length	NOV79a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O60885	Bromodomain-containing protein 4 (HUNK1 protein) - Homo sapiens (Human), 1362 aa.	1..1390 1..1362	1357/1391 (97%) 1360/1391 (97%)	0.0
Q9ESU6	CELL PROLIFERATION RELATED PROTEIN CAP - Mus musculus (Mouse), 1400 aa.	1..1390 1..1400	1318/1400 (94%) 1338/1400 (95%)	0.0
AAL67833	BROMODOMAIN-CONTAINING PROTEIN BRD4 LONG VARIANT - Mus musculus (Mouse), 1400 aa.	1..1390 1..1400	1318/1400 (94%) 1338/1400 (95%)	0.0
O60433	R31546_1 - Homo sapiens (Human), 731 aa (fragment).	1..719 12..730	719/719 (100%) 719/719 (100%)	0.0
AAL67834	BROMODOMAIN-CONTAINING PROTEIN BRD4 SHORT VARIANT - Mus musculus (Mouse), 723 aa.	1..719 1..720	694/720 (96%) 700/720 (96%)	0.0

PFam analysis predicts that the NOV79a protein contains the domains shown in the

5   Table 79E.

328

<table>
<tr><td colspan="4" align="center">**Table 79E. Domain Analysis of NOV79a**</td></tr>
</table>

Pfam Domain	NOV79a Match Region	Identities/ Similarities for the Matched Region	Expect Value
bromodomain: domain 1 of 2	63..152	42/92 (46%) 82/92 (89%)	8.6e-45
bromodomain: domain 2 of 2	356..445	40/92 (43%) 81/92 (88%)	3e-40

Example 80.

The NOV80 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 80A.

<table>
<tr><td colspan="4" align="center">**Table 80A. NOV80 Sequence Analysis**</td></tr>
<tr><td></td><td>SEQ ID NO: 225</td><td colspan="2">1776 bp</td></tr>
<tr><td>NOV80a,
CG59572-01 DNA Sequence</td><td colspan="3">TGGTTCGTTTATTCCTGGGGTTGTCATATCATGGCTTATAATGACACAGACAGAAACC
AGACTGAGAAGCTCCTAAAAAGAGTACGAGAACTGGAGCAAGAGGTGCAAAGACTTAA
AAAGGAACAGGCCAAAAATAAGGAGGACTCAAACATTAGAGAAAATTCAGCAGGAGCT
GGAAAAACTAAGCGTGCATTTGATTTCAGTGCTCATGGCCGAAGCACGTAGCCCTAA
GAATAGCCTATATGGGCTGGGGATACCAGGGCTTTGCTAGTCAGGAAAACACAAATAA
TACCATTGAAGAGAAACTGTTTGAAGCTCTAACCAAGACTCGACTAGTAGAAAGCAGA
CAGACATCCAACTATCACCGATGTGGGAGAACAGATAAAGGAGTTAGTGCCTTTGGAC
AGGTGATCTCACTTGACCTTCGCTCTCAGTTTCCAAGGGGCAGGGATTCCGAGGACTT
TAATGTAAAAGAGGAGGCTAATGCTGCTGCTGAAGAGATCCGTTATACCCACATTCTC
AATCGGGTACTCCCTCCAGACATCCGTATATTGGCCTGGGCCCCTGTAGAACCAAGCT
TCAGTGCTAGGTTCAGCTGCCTTGAGCGGACTTACCGCTATTTTTTCCCTCGTGCTGA
TTTAGATATTGTAACCATGGATTATGCAGCTCAGAAGTATGTTGGCACCCATGATTTC
AGGAACTTGTGTAAAATGGATGTAGCCAACGGTGTGATTAATTTTCAGAGGACTATTC
TATCTGCTCAAGTACAGCTAGTGGGCCAGAGCCCAGGTGAGGGGAGATGGCAAGAACC
TTTCCAGTTATGTCAGTTTGAAGTGACTGGCCAGGCATTCCTTTATCATCAAGTCCGA
TGTATGATGGCTATCCTCTTTCTGATTGGCCAAGGAATGGAGAAGCCAGAGATTATTG
ATGAGCTGCTGAATATAGAGAAAAATCCCCAAAAGCCTCAATATAGTATGGCTGTAGA
ATTTCCTCTAGTCTTATATGACTGTAAGTTTGAAAATGTCAAGTGGATCTATGACCAG
GAGGCTCAGGAGTTCAATATTACCCACCTACAACAACTGTGGGCTAATCATGCTGTCA
AAACTCACATGTTGTATAGTATGCTACAAGGACTGGACACTGTTCCAGTACCCTGTGG
AATAGGACCAAAGATGGATGGAATGACAGAATGGGGAAATGTTAAGCCCTCTGTCATA
AAGCAGACCAGTGCCTTTGTAGAAGGAGTGAAGATGCGCACATATAAGCCCCTCATGG
ACCGTCCTAAATGCCAAGGACTGGAATCCCGGATCCAGCATTTTGTACGTAGGGGACG
AATTGAGCACCCACATTTATTCCATGAGGAAGAAACAAAAGCCAAAAGGGACTGTAAT
GACACACTAGAGGAAGAGAATACTAATTTGGAGACACCAACGAAGAGGGTCTGTGTTG
ACACAGAAATTAAAAGCATCATT**TAA**CCATAGACAATTTGCCAGGATCTAGGAACCAC
CTAATGGTAGGTGGACAGAAAGGAAAAAAAAAAAAATTTACTTGCAAGTACTAGGAA
TTCAGATGATCAGCTCTTAAAAAAAAAAAAAAAAAAGCAAAAAGACTAAAGCCCTATTAAG
GAAGTTATTGCTTTAATAAGAAATTTCAAATATTCTCTTATCCCGGTCCAAAAGGATT
AAGCGATTAAAGAACGTAAAATGGAGATGTATTTACATACACCTGGAAACCTGTGCCT
TGTATTCAAATTCATTAAAGCCTAATCCTGCAAGAA</td></tr>
<tr><td></td><td>ORF Start: ATG at 31</td><td colspan="2">ORF Stop: TAA at 1474</td></tr>
<tr><td></td><td>SEQ ID NO: 226</td><td>481 aa</td><td>MW at 55646.8kD</td></tr>
<tr><td>NOV80a,
CG59572-01 Protein Sequence</td><td colspan="3">MAYNDTDRNQTEKLLKRVRELEQEVQRLKKEQAKNKEDSNIRENSAGAGKTKRAFDFS
AHGRRHVALRIAYMGWGYQGFASQENTNNTIEEKLFEALTKTRLVESRQTSNYHRCGR
TDKGVSAFGQVISLDLRSQFPRGRDSEDFNVKEEANAAAEEIRYTHILNRVLPPDIRI
LAWAPVEPSFSARFSCLERTYRYFFPRADLDIVTMDYAAQKYVGTHDFRNLCKMDVAN
GVINFQRTILSAQVQLVGQSPGEGRWQEPFQLCQFEVTGQAFLYHQVRCMMAILFLIG
QGMEKPEIIDELLNIEKNPQKPQYSMAVEFPLVLYDCKFENVKWIYDQEAQEFNITHL
QQLWANHAVKTHMLYSMLQGLDTVPVPCGIGPKMDGMTEWGNVKPSVIKQTSAFVEGV
KMRTYKPLMDRPKCQGLESRIQHFVRRGRIEHPHLFHEEETKAKRDCNDTLEEENTNL
ETPTKRVCVDTEIKSII</td></tr>
<tr><td></td><td>SEQ ID NO: 227</td><td colspan="2">1508 bp</td></tr>
</table>

329

NOV80b, CG59572-02 DNA Sequence	CATGGCTTATAATGACACAGACAGAAACCAGACTGAGAAGCTCCTAAAAAGAGTACGA GAACTGGAGCAAGAGGTGCAAAGACTTAAAAAGGAACAGGCCAAAAATAAGGAGGACT CAAACATTAGAGAAAATTCAGCAGGAGCTGGAAAAACTAAGCGTGCATTTGATTTCAG TGCTCATGGCCGAAGACACGTAGCCCTAAGAATAGCCTATATGGGCTGGGGATACCAG GGCTTTGCTAGTCAGGAAAACACAAATAATACCATTGAAGAGAAACTGTTTGAAGCTC TAACCAAGACTCGACTAGTAGAAAGCAGACAGACATCCAACTATCACCGATGTGGGAG AACAGATAAAGGAGTTAGTGCCTTTGGACAGGTGATCTCACTTGACCTTCGCTCTCAG TTTCCAAGGGGCAGGGATTCCGAGGACTTTAATGTAAAAGAGGAGGCTAATGCTGCTG CTGAAGAGATCCGTTATACCCACATTCTCAATCGGGTACTCCCTCCAGACATCCGTAT ATTGGCCTGGGCCCCTGTAGAACCAAGCTTCAGTGCTAGGTTCAGCTGCCTTGAGCGG ACTTACCGCTATTTTTTCCCTCGTGCTGATTTAGATATTGTAACCATGGATTATGCAG CTCAGAAGTATGTTGGCACCCATGATTTCAGGAACTTGTGTAAAATGGATGTAGCCAA CGGTGTGATTAATTTTCAGAGGACTATTCTATCTGCTCAAGTACAGCTAGTGGGCCAG AGCCCAGGTGAGGGGAGATGGCAAGAACCTTTCCAGTTATGTCAGTTTGAAGTGACTG GCCAGGCATTCCTTTATCATCAAGTCCGATGTATGATGGCTATCCTCTTTCTGATTGG CCAAGGAATGGAGAAGCCAGAGATTATTGATGAGCTGCTGAATATAGAGAAAAATCCC CAAAAGCCTCAATATAGTATGGCTGTAGAATTTCCTCTAGTCTTATATGACTGTAAGT TTGAAAATGTCAAGTGGATCTATGACCAGGAGGCTCAGGAGTTCAATATTACCCACCT ACAACAACTGTGGGCTAATCATGCTGTCAAAACTCACATGTTGTATAGTATGCTACAA GGACTGGCACACTGTTCCAGTACCCTGTGGAATAGGACCAAAGATGGATGGAATGACAG AATGGGGAAATGTTAAGCCCTCTGTCATAAAGCAGACCAGTGCCTTTGTAGAAGGAGT GAAGATGCGCACATATAAGCCCCTCATGGACCGTCCTAAATGCCAAGGACTGGAATCC CGGATCCAGCATTTTGTACGTAGGGGACGAATTGAGCACCCACATTTATTCCATGAGG AAGAAACAAAAGCCAAAAGGGACTGTAATGACACACTAGAGGAAGAGAATACTAATTT GGAGACACCAACGAAGAGGGTCTGTGTTGACACAGAAATTAAAAGTATCATTTAACCA TAGACAATTTGCCAGGATCTAGGAACCACCTAATGGTAGGTGGACAGAAAAGGAAAAA
	ORF Start: ATG at 2	ORF Stop: TAA at 1445
	SEQ ID NO: 228	481 aa	MW at 55646.8kD
NOV80b, CG59572-02 Protein Sequence	MAYNDTDRNQTEKLLKRVRELEQEVQRLKKEQAKNKEDSNIRENSAGAGKTKRAFDFS AHGRRHVALRIAYMGWGYQGFASQENTNNTIEEKLFEALTKTRLVESRQTSNYHRCGR TDKGVSAFGQVISLDLRSQFPRGRDSEDFNVKEEANAAAEEIRYTHILNRVLPPDIRI LAWAPVEPSFSARFSCLERTYRYFFPRADLDIVTMDYAAQKYVGTHDFRNLCKMDVAN GVINFQRTILSAQVQLVGQSPGEGRWQEPFQLCQFEVTGQAFLYHQVRCMMAILFLIG QGMEKPEIIDELLNIEKNPQKPQYSMAVEFPLVLYDCKFENVKWIYDQEAQEFNITHL QQLWANHAVKTHMLYSMLQGLDTVPVPCGIGPKMDGMTEWGNVKPSVIKQTSAFVEGV KMRTYKPLMDRPKCQGLESRIQHFVRRGRIEHPHLFHEEETKAKRDCNDTLEEENTNL ETPTKRVCVDTEIKSII

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 80B.

Table 80B. Comparison of NOV80a against NOV80b.		
Protein Sequence	NOV80a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV80b	1..481 1..481	459/481 (95%) 459/481 (95%)

Further analysis of the NOV80a protein yielded the following properties shown in Table 80C.

Table 80C. Protein Sequence Properties NOV80a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0142 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV80a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 80D.

Table 80D. Geneseq Results for NOV80a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV80a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM79457	Human protein SEQ ID NO 3103 - Homo sapiens, 490 aa. [WO200157190-A2, 09-AUG-2001]	1..481 10..490	478/481 (99%) 480/481 (99%)	0.0
AAM78473	Human protein SEQ ID NO 1135 - Homo sapiens, 481 aa. [WO200157190-A2, 09-AUG-2001]	1..481 1..481	478/481 (99%) 480/481 (99%)	0.0
AAG64907	Human depressed growth rate protein DEG1 - Homo sapiens, 248 aa. [CN1296014-A, 23-MAY-2001]	209..431 1..223	223/223 (100%) 223/223 (100%)	e-132
AAG02637	Human secreted protein, SEQ ID NO: 6718 - Homo sapiens, 96 aa. [EP1033401-A2, 06-SEP-2000]	361..456 1..96	96/96 (100%) 96/96 (100%)	5e-53
AAB96592	Putative P. abyssi pseudourydilate synthase I - Pyrococcus abyssi, 263 aa. [FR2792651-A1, 27-OCT-2000]	65..367 3..261	79/305 (25%) 140/305 (45%)	4e-16

5   In a BLAST search of public sequence databases, the NOV80a protein was found to have homology to the proteins shown in the BLASTP data in Table 80E.

Table 80E. Public BLASTP Results for NOV80a

Protein Accession Number	Protein/Organism/Length	NOV80a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BZE2	FKSG32 - Homo sapiens (Human), 481 aa.	1..481 1..481	481/481 (100%) 481/481 (100%)	0.0
Q96J23	HYPOTHETICAL 55.6 KDA PROTEIN - Homo sapiens (Human), 481 aa.	1..481 1..481	478/481 (99%) 480/481 (99%)	0.0
Q96NB4	CDNA FLJ31140 FIS, CLONE IMR322001218, HIGHLY SIMILAR TO MUS MUSCULUS PSEUDOURIDINE SYNTHASE 3 (PUS3) MRNA - Homo sapiens (Human), 481 aa.	1..481 1..481	478/481 (99%) 479/481 (99%)	0.0
Q9JI38	PSEUDOURIDINE SYNTHASE 3 - Mus musculus (Mouse), 481 aa.	5..480 4..480	407/479 (84%) 434/479 (89%)	0.0
Q9D0F7	2610020J05RIK PROTEIN - Mus musculus (Mouse), 316 aa.	5..314 4..315	276/312 (88%) 291/312 (92%)	e-158

PFam analysis predicts that the NOV80a protein contains the domains shown in the Table 80F.

Table 80F. Domain Analysis of NOV80a

Pfam Domain	NOV80a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PseudoU_synth_1: domain 1 of 1	88..307	70/249 (28%) 176/249 (71%)	4.7e-57

Example 81.

The NOV81 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 81A.

Table 81A. NOV81 Sequence Analysis

	SEQ ID NO: 229	3080 bp
NOV81a, CG59522-01 DNA Sequence	TTCCAGCCGGCAGGATGGAGGACGAGGAAGGCCCTGAGTATGGCAAACCTGACTTTGT GCTTTTGGACCAAGTGACCATGGAGGACTTCATGAGGAACCTGCAGCTCAGGTTCGAG AAGGGCCGCATCTACACCTACATCGGTGAGGTGCTGGTGTCCGTGAACCCCTACCAGG AGCTGCCCCTGTATGGGCCTGAGGCCATCGCCAGGTACCAGGGCCGTGAGCTCTATGA GCGGCCACCCCATCTCTATGCTGTGGCCAACGCCGCCTACAAGGCAATGAAGCACCGG	

| | TCCAGGGACACCTGCATCGTCATCTCAGGGGAGAGTGGGGCAGGGAAGACAGAAGCCA GTAAGCACATCATGCAGTACATCGCTGCTGTCACCAATCCAAGCCAGAGGGCTGAGGT GGAGAGGGTCAAGGACGTGCTGCTCAAGTCCACCTGTGTGCTGGAGGCCTTTGGCAAT GCCCGCACCAACCGCAATCACAACTCCAGCCGCTTTGGCAAGTACATGGACATCAACT TTGACTTCAAGGGGGGACCCGATCGGAGGACACATCCACAGCTACCTACTGGAGAAGTC TCGGGTCCTCAAGCAGCACGTGGGTGAAAGAAACTTCCACGCCTTCTACCAATTGCTG AGAGGCAGTGAGGACAAGCAGCTGCATGAACTGCACTTGGAGAGAAACCCTGCTGTAT ACAATTTCACACACCAGGGAGCAGGACTCAACATGACTGTGAGTGATGAGCAGAGCCA CCAGGCAGTGACCGAGGCCATGAGGGTCATCGGCTTCAGTCCTGAAGAGGTGGAGTCT GTGCATCGCATCCTGGCTGCCATATTGCACCTGGGAAACATCGAGTTTGTGGAGACGG AGGAGGGTGGGCTGCAGAAGGAGGGCCTGGCAGTGGCCGAGGAGGCACTGGTGGACCA TGTGGCTGAGCTGACGGCCACACCCCGGGACCTCGTGCTCCGCTCCCTGCTGGCTCGC ACAGTTGCCTCGGGAGGCAGGGAACTCATAGAGAAGGGCCACACTGCAGCTGAGGCCA GCTATGCCCGGGATGCCTGTGCCAAGGCAGTGTACCAGCGGCTGTTTGAGTGGGTGGT GAACAGGATCAACAGTGTCATGGAACCCCGGGGCCGGGATCCTCGGCGTGATGGCAAG GACACAGTCATTGGCGTGCTGGACATCTATGGCTTCGAGGTGTTTCCCGTCAACAGTT TCGAGCAGTTCTGCATCAACTACTGCAACGAGAAGCTGCAGCAGCTATTCATCCAGCT CATCCTGAAGCAGGAACAGGAAGAGTACGAGCGCGAGGGCATCACCTGGCAGAGCGTT GAGTATTTCAACAACGCCACCATTGTGGATCTGGTGGAGCGGCCCCACCGTGGCATCC TGGCCGTGCTGGACGAGGCCTGCAGCTCTGCTGGCACCATCACTGACCGAATCTTCCT GCAGACCCTGGACATGCACCACCGCCATCACCTACACTACACCAGCCGCCAGCTCTGC CCCACAGACAAGACCATGGAGTTTGGCCGAGACTTCCGGATCAAGCACTATGCAGGGG ACGTCACGTACTCCGTGGAAGGCTTCATCGACAAGAACAGAGATTTCCTCTTCCAGGA CTTCAAGCGGCTGCTGTACAACAGCACGGACCCCACTCTACGGGCCATGTGGCCGGAC GGGCAGCAGGACATCACAGAGGTGACCAAGCGCCCCCTGACGGCTGGCACACTCTTCA AGAACTCCATGGTGGCCCTGGTGGAGAACCTTGCCTCCAAGGAGCCCTTCTACGTCCG CTGCATCAAGCCCAATGAGGACAAGGTAGCTGGGAAGCTGGATGAGAACCACTGTCGC CACCAGGTCGCATACCTGGGGCTGCTGGAGAATGTGAGGGTCCGCAGGGCTGGCTTCG CTTCCCGCCAGCCCTACTCTCGATTCCTGCTCAGGTACAAGATGACCTGTGAATACAC ATGGCCCAACCACCTGCTGGGCTCCGACAAGGCAGCCGTGAGCGCTCTCCTGGAGCAG CACGGGCTGCAGGGGGACGTGGCCTTTGGCCACAGCAAGCTGTTCATCCGCTCACCCC GGACACTGGTCACACTGGAGCAGAGCCGAGCCCGCCTCATCCCCATCATTGTGCTGCT ATTGCAGAAGGCATGGCGGGGCACCTTGGCGAGGTGGCGCTGCCGGAGGCTGAGGGCT ATCTACACCATCATGCGCTGGTTCCGGAGACACAAGGTGCGGGCTCACCTGGCTGAGC TGCAGCGGCGATTCCAGGCTGCAAGGCAGCCGCCACTCTACGGGCGTGACCTTGTGTG GCCGCTGCCCCCTGCTGTGCTGCAGCCCTTCCAGGACACCTGCCACGCACTCTTCTGC AGGTGGCGGGCCCGGCAGCTGGTGAAGAACATCCCCCCTTCAGACATGCCCCAGATCA AGGCCAAGGTGGCCGCCATGGGGGGCCCTGCAAGGGCTTCGTCAGGACTGGGGCTGCCG ACGGGCCTGGGCCCGAGACTACCTGTCCTCTGCCACTGACAATCCCACAGCATCAAGC CTGTTTGCTCAGCGACTAAAGACACTTCAGGACAAAGATGGCTTCGGGGCTGTGCTCT TTTCAAGCCATGTCCGCAAGGTGAACCGCTTCCACAAGATCCGGAACCGGGCCCTCCT GCTCACAGACCAGCACCTCTACAAGCTGGACCCTGACCGGCAGTACCGGGTGATGCGG GCCGTGCCCCTTGAGGCGGTGACGGGGCTGAGCGTGACCAGCGGAGGAGACCAGCTGG TGGTGCTGCACGCCCGCGGCCAGGACGACCTCGTGGTGTGCCTGCACCGCTCCCGGCC GCCATTGGACAACCGCGTTGGGGAGCTGGTGGGCGTGCTGGCCCGCACACTGCCGCAGG GAGGGCCGCACCCTGGAGGTTCGCGTCTCCGACTGCATCCCACTAAGCCATCGCGGGG TCCGGCGCCTCATCTCCGTGGAGCCCAGGCCGGAGCAGCCAGAGCCCGATTCCGCTG CGCTCGCGGCTCCTTCACCCTGCTCTGCCCCAGCCGC**TGA**GCGCCCGCACCCGCCGCA CCCCGA |

	ORF Start: ATG at 15	ORF Stop: TGA at 3054	
	SEQ ID NO: 230	1013 aa	MW at 116044.5kD
NOV81a, CG59522-01 Protein Sequence	MEDEEGPEYGKPDFVLLDQVTMEDFMRNLQLRFEKGRIYTYIGEVLVSVNPYQELPLY GPEAIARYQGRELYERPPHLYAVANAAYKAMKHRSRDTCIVISGESGAGKTEASKHIM QYIAAVTNPSQRAEVERVKDVLLKSTCVLEAFGNARTNRNHNSSRFGKYMDINFDFKG DPIGGHIHSYLLEKSRVLKQHVGERNFHAFYQLLRGSEDKQLHELHLERNPAVYNFTH QGAGLNMTVSDEQSHQAVTEAMRVIGFSPEEVESVHRILAAILHLGNIEFVETEEGGL QKEGLAVAEEALVDHVAELTATPRDLVLRSLLARTVASGGRELIEKGHTAAEASYARD ACAKAVYQRLFEWVVNRINSVMEPRGRDPRRDGKDTVIGVLDIYGFEVFPVNSFEQFC INYCNEKLQQLFIQLILKQEQEEYEREGITWQSVEYFNNATIVDLVERPHRGILAVLD EACSSAGTITDRIFLQTLDMHHRHHLHYTSRQLCPTDKTMEFGRDFRIKHYAGDVTYS VEGFIDKNRDFLFQDFKRLLYNSTDPTLRAMWPDGQQDITEVTKRPLTAGTLFKNSMV ALVENLASKEPFYVRCIKPNEDKVAGKLDENHCRHQVAYLGLLENVRVRRAGFASRQP YSRFLLRYKMTCEYTWPNHLLGSDKAAVSALLEQHGLQGDVAFGHSKLFIRSPRTLVT LEQSRARLIPIIVLLLQKAWRGTLARWRCRRLRAIYTIMRWFRRHKVRAHLAELQRRF QAARQPPLYGRDLVWPLPPAVLQPFQDTCHALFCRWRARQLVKNIPPSDMPQIKAKVA AMGALQGLRQDWGCRRAWARDYLSSATDNPTASSLFAQRLKTLQDKDGFGAVLFSSHV RKVNRFHKIRNRALLLTDQHLYKLDPDRQYRVMRAVPLEAVTGLSVTSGGDQLVVLHA RGQDDLVVCLHRSRPPLDNRVGELVGVLAAHCRREGRTLEVRVSDCIPLSHRGVRRLI SVEPRPEQPEPDFRCARGSFTLLWPSR

333

Further analysis of the NOV81a protein yielded the following properties shown in Table 81B.

Table 81B. Protein Sequence Properties NOV81a	
PSort analysis:	0.8800 probability located in nucleus; 0.3902 probability located in microbody (peroxisome); 0.2210 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV81a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 81C.

Table 81C. Geneseq Results for NOV81a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV81a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU23125	Novel human enzyme polypeptide #211 - Homo sapiens, 1026 aa. [WO200155301-A2, 02-AUG-2001]	1..1013 9..1026	1009/1018 (99%) 1011/1018 (99%)	0.0
AAU23128	Novel human enzyme polypeptide #214 - Homo sapiens, 909 aa. [WO200155301-A2, 02-AUG-2001]	1..853 9..866	851/858 (99%) 851/858 (99%)	0.0
AAM80123	Human protein SEQ ID NO 3769 - Homo sapiens, 764 aa. [WO200157190-A2, 09-AUG-2001]	243..1011 1..762	438/769 (56%) 570/769 (73%)	0.0
AAM79139	Human protein SEQ ID NO 1801 - Homo sapiens, 753 aa. [WO200157190-A2, 09-AUG-2001]	254..1011 1..751	434/758 (57%) 564/758 (74%)	0.0
AAM39991	Human polypeptide SEQ ID NO 3136 - Homo sapiens, 1063 aa. [WO200153312-A1, 26-JUL-2001]	10..933 47..986	410/966 (42%) 556/966 (57%)	0.0

In a BLAST search of public sequence databases, the NOV81a protein was found to have homology to the proteins shown in the BLASTP data in Table 81D.

Table 81D. Public BLASTP Results for NOV81a				
Protein Accession Number	Protein/Organism/Length	NOV81a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q63357	MYOSIN I - Rattus norvegicus (Rat), 1006 aa.	1..1011 1..1004	606/1011 (59%) 780/1011 (76%)	0.0
A53933	myosin I myr 4 - rat, 1006 aa.	1..1011 1..1004	604/1011 (59%) 778/1011 (76%)	0.0
Q96RI6	UNCONVENTIONAL MYOSIN 1G VALINE FORM - Homo sapiens (Human), 633 aa (fragment).	33..646 1..619	612/619 (98%) 612/619 (98%)	0.0
Q96RI5	UNCONVENTIONAL MYOSIN 1G METHONINE FORM - Homo sapiens (Human), 633 aa (fragment).	33..646 1..619	611/619 (98%) 612/619 (98%)	0.0
Q23978	Myosin IA (MIA) (Brush border myosin IA) (BBMIA) - Drosophila melanogaster (Fruit fly), 1011 aa.	8..1012 6..1007	503/1017 (49%) 686/1017 (66%)	0.0

PFam analysis predicts that the NOV81a protein contains the domains shown in the Table 81E.

Table 81E. Domain Analysis of NOV81a			
Pfam Domain	NOV81a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PRK: domain 1 of 1	97..109	8/13 (62%) 10/13 (77%)	3.7
Vir_DNA_binding: domain 1 of 1	575..592	5/18 (28%) 14/18 (78%)	8.2
myosin_head: domain 1 of 1	11..689	305/747 (41%) 531/747 (71%)	8.1e-288

335

Example 82.

The NOV82 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 82A.

Table 82A. NOV82 Sequence Analysis		
	SEQ ID NO: 231	1066 bp
NOV82a, CG59520-01 DNA Sequence	GAACGAATGGGAAACCAGAAATCAGATATTTATGCCCAAGCAAAGCAGGATTTCGTTC AGCACTACTCCCAGATCGTTAGGGTGCTGACTGAGGATGAGATGGGGCACCCAGAGAC AGGAGATGCTACTGCCCGGCTCAAGGAGGTCCTGGAGTACAATGCCATTGGAGGCAAG TATCACCGAGGTTTGATGGTGCTAGTAGCGTTCCGGGAGCTGGTGGAGCCGAGGAAAC TGGATGCTGATAGTCTCCAGTGGGCACCGACTGTGGGCTGGTATGCGCAACTGCTGCA AGCTTTCTTCCTGGTGGCAGATGACATTATGGATTCATCCCTTACCTGCCAGGGACAG ATCTCCTGGTATCAGAAGCTGGGCATGGGTTTGGATGCCATCAATGATGCTATCCTTC TGGAAGCATGTATCTACTGCCTGCTGAAGCTGTATTGCCGGGAGCAGCCCTATTACCT GAACCTGATGGAGCTCTTCCAGCAGAATTCTTATCAGACTGAGATTGGGCAGACCCTC GACCTCATCACAACCCCCCAGGGCAATGTGGATCTTCGCAGATGCACCGAAAAAAGGC ACAAATCTGTTGTCAAGTACAAGACAGCTTTCTACTCCTTCTACCTTCCTGTAGCTGC AGCCATGTACATGTCAAGAATGGATGACAAGAAGGAGCACACCAGTGCCAAGAAGATC CTGCTGGAGATTCAAGAGTTCTTTCAGATTCAGGATGATTACCTTGACTTCTCTGGGG ACCCCAGTGTGACTGGCAGAGTTGGCAATGACTTCCAGGACAACAAATGCAGCTGGCT GGTGGTTCAGTGTCTGCTACAGGCCACTCCAGAACAGTACCAGATCCTGAAGGAAAAT TACAGGCAGAAGGAGGCCGAGAAGGTGGCCCGGGTGAAGGCACTATACGAGGAGCTGG ATCTGCCAGCCGTGTTCTTGCAGTATGAGAAAGACAGTTACAGCCACGTTATGGGTCT CATCGAACAGTACGCAGAGCCCCTGCCCCCAGCCATCTTTCTGGGGCTTGTGCACAAA ATCTACAAGTGGAAAAAGTGAC
	ORF Start: ATG at 7	ORF Stop: TGA at 1063
	SEQ ID NO: 232	352 aa	MW at 40740.3kD
NOV82a, CG59520-01 Protein Sequence	MGNQKSDIYAQAKQDFVQHYSQIVRVLTEDEMGHPETGDATARLKEVLEYNAIGGKYH RGLMVLVAFRELVEPRKLDADSLQWAPTVGWYAQLLQAFFLVADDIMDSSLTCQGQIS WYQKLGMGLDAINDAILLEACIYCLLKLYCREQPYYLNLMELFQQNSYQTEIGQTLDL ITTPQGNVDLRRCTEKRHKSVVKYKTAFYSFYLPVAAAMYMSRMDDKKEHTSAKKILL EIQEFFQIQDDYLDFSGDPSVTGRVGNDFQDNKCSWLVVQCLLQATPEQYQILKENYR QKEAEKVARVKALYEELDLPAVFLQYEKDSYSHVMGLIEQYAEPLPPAIFLGLVHKIY KWKK

Further analysis of the NOV82a protein yielded the following properties shown in Table 82B.

Table 82B. Protein Sequence Properties NOV82a	
PSort analysis:	0.4066 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV82a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 82C.

336

Table 82C. Geneseq Results for NOV82a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV82a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG29733	Arabidopsis thaliana protein fragment SEQ ID NO: 35427 - Arabidopsis thaliana, 342 aa. [EP1033405-A2, 06-SEP-2000]	10..352 2..342	147/343 (42%) 219/343 (62%)	7e-75
AAG29732	Arabidopsis thaliana protein fragment SEQ ID NO: 35426 - Arabidopsis thaliana, 349 aa. [EP1033405-A2, 06-SEP-2000]	10..352 9..349	147/343 (42%) 219/343 (62%)	7e-75
AAG29734	Arabidopsis thaliana protein fragment SEQ ID NO: 35428 - Arabidopsis thaliana, 305 aa. [EP1033405-A2, 06-SEP-2000]	47..352 1..305	138/306 (45%) 204/306 (66%)	4e-73
AAY43635	Amino acid sequence of the farnesyl pyrophosphate synthase enzyme - Phaffia rhodozyma, 355 aa. [EP955363-A2, 10-NOV-1999]	12..352 11..355	145/346 (41%) 208/346 (59%)	4e-69
AAB48971	Sunflower seedling farnesyl pyrophosphate synthase (FPS) - Helianthus annuus, 341 aa. [EP1063297-A1, 27-DEC-2000]	13..352 6..341	138/343 (40%) 204/343 (59%)	3e-64

In a BLAST search of public sequence databases, the NOV82a protein was found to have homology to the proteins shown in the BLASTP data in Table 82D.

337

Table 82D. Public BLASTP Results for NOV82a				
Protein Accession Number	Protein/Organism/Length	NOV82a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96G29	FARNESYL DIPHOSPHATE SYNTHASE (FARNESYL PYROPHOSPHATE SYNTHETASE, DIMETHYLALLYLTRANSTRA NSFERASE, GERANYLTRANSTRANSFERA SE) - Homo sapiens (Human), 419 aa.	2..352 69..419	291/351 (82%) 317/351 (89%)	e-168
P14324	Farnesyl pyrophosphate synthetase (FPP synthetase) (FPS) (Farnesyl diphosphate synthetase) [Includes: Dimethylallyltransferase (EC 2.5.1.1); Geranyltranstransferase (EC 2.5.1.10)] - Homo sapiens (Human), 353 aa.	2..352 3..353	291/351 (82%) 317/351 (89%)	e-168
A35726	farnesyl-pyrophosphate synthetase - human, 353 aa.	2..352 3..353	290/351 (82%) 316/351 (89%)	e-168
AAL58886	FARNESYL DIPHOSPHATE SYNTHASE - Bos taurus (Bovine), 353 aa.	2..352 3..353	270/351 (76%) 308/351 (86%)	e-157
Q14329	FARNESYL PYROPHOSPHATE SYNTHETASE LIKE-4 PROTEIN - Homo sapiens (Human), 348 aa.	6..352 2..348	268/347 (77%) 295/347 (84%)	e-150

PFam analysis predicts that the NOV82a protein contains the domains shown in the Table 82E.

Table 82E. Domain Analysis of NOV82a			
Pfam Domain	NOV82a Match Region	Identities/ Similarities for the Matched Region	Expect Value
polyprenyl_synt: domain 1 of 1	43..315	82/285 (29%) 237/285 (83%)	6.3e-91

338

Example 83.

The NOV83 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 83A.

Table 83A. NOV83 Sequence Analysis		
	SEQ ID NO: 233	411 bp
NOV83a, CG59758-01 DNA Sequence	TGCCTACCCCGAGACTGCTGCTGTTCGGAGACCTGCAGGTGAATGCCCCATCACCATG TCTGACCTGGAGGCAAAACCTTCAACTGAGCATTTGGGGGATAAGATAAAAGATGAAG ATATTAAACTCAGGGTTATTGGACAGGATAGCAGTGAGATTCATTTCAAAGTGAAAAT GACAACACCTCTCAAGAAACTCAAGAAATCGTACTGTCAGAGACAGGGCGTTCCAGTG AATTCCCTCAGGTTTCTCTTTGAAGGTCAGAGAATTGCTGATAATCATACTCCAGAAG AACTGGGAATGGAGGAAGAAGATGTGATTGAGGTTTATCAGGAACAAATCGGAGGTCA TTCAACAGTTTAGACATTCTTTTTTTTTTTCCTTTTCCCTCAATCCTTTTTTATTTTT TTAAA	
	ORF Start: ATG at 56	ORF Stop: TAG at 359
	SEQ ID NO: 234	101 aa | MW at 11526.0kD
NOV83a, CG59758-01 Protein Sequence	MSDLEAKPSTEHLGDKIKDEDIKLRVIGQDSSEIHFKVKMTTPLKKLKKSYCQRQGVP VNSLRFLFEGQRIADNHTPEELGMEEEDVIEVYQEQIGGHSTV	
	SEQ ID NO: 235	658 bp
NOV83b, CG59758-02 DNA Sequence	CTACCCCGAGACTGCTGCTGTTCGGAGACCTGCAGGTGAATGCCCCATCACCATGTCT GACCTGGAGGCAAAACCTTCAACTGAGCATTTGGGGGATAAGATAAAAGATGAAGATA TTAAACTCAGGGTTATTGGACAGGATAGCAGTGAGATTCATTTCAAAGTGAAAATGAC AACACCTCTCAAGAAACTCAAGAAATCGTACTGTCAGAGACAGGGCGTTCCAGTGAAT TCCCTCAGGTTTCTCTTTGAAGGTCAGAGAATTGCTGATAATCATACTCCAGAAGAAC TGGGAATGGAGGAAGAAGATGTGATTGAGGTTTATCAGGAACAAATCGGAGGTCATTC AACAGTTTAGACAATCGGAGGTCATTCAACAGTTTAGACAATCGGAGGTCATTCAACA GTTTAGACAATCGGAGGTCATTCAACAGTTTAGACAATCGGAGGTCATTCAACAGTTT AGACAATCGGAGGTCATTCAACAGTTTAGACAATCGGAGGTCATTCAACAGTTTAGAC AATCGGAGGTCATTCAACAGTTTAGACAATCGGAGGTCATTCAACAGTTTAGACAATC GGAGGTCATTCAACAGTTTAGACAATCGGAGGTCATTCAACAGTTTAGACAATCGGAG GTCATTCAACAGTTTAGACA	
	ORF Start: ATG at 53	ORF Stop: TAG at 356
	SEQ ID NO: 236	101 aa | MW at 11526.0kD
NOV83b, CG59758-02 Protein Sequence	MSDLEAKPSTEHLGDKIKDEDIKLRVIGQDSSEIHFKVKMTTPLKKLKKSYCQRQGVP VNSLRFLFEGQRIADNHTPEELGMEEEDVIEVYQEQIGGHSTV	

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 83B.

Table 83B. Comparison of NOV83a against NOV83b.		
Protein Sequence	NOV83a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV83b	1..101 1..101	101/101 (100%) 101/101 (100%)

Further analysis of the NOV83a protein yielded the following properties shown in Table 83C.

339

Table 83C. Protein Sequence Properties NOV83a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV83a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 83D.

Table 83D. Geneseq Results for NOV83a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV83a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM79976	Human protein SEQ ID NO 3622 - Homo sapiens, 125 aa. [WO200157190-A2, 09-AUG-2001]	1..101 25..125	100/101 (99%) 100/101 (99%)	1e-52
AAM78992	Human protein SEQ ID NO 1654 - Homo sapiens, 101 aa. [WO200157190-A2, 09-AUG-2001]	1..101 1..101	100/101 (99%) 100/101 (99%)	1e-52
AAY49967	Human sentrin protein sequence - Homo sapiens, 101 aa. [US5985664-A, 16-NOV-1999]	1..101 1..101	89/101 (88%) 94/101 (92%)	2e-45
AAW87984	Ubiquitin-like domain of the protein SUMO1 - Mammalia, 101 aa. [WO9857978-A1, 23-DEC-1998]	1..101 1..101	89/101 (88%) 94/101 (92%)	2e-45
AAW60079	Homo sapiens sentrin-1 polypeptide - Homo sapiens, 101 aa. [WO9820038-A1, 14-MAY-1998]	1..101 1..101	89/101 (88%) 94/101 (92%)	2e-45

5   In a BLAST search of public sequence databases, the NOV83a protein was found to have homology to the proteins shown in the BLASTP data in Table 83E.

340

<table>
<tr><th colspan="5">Table 83E. Public BLASTP Results for NOV83a</th></tr>
<tr>
<th>Protein Accession Number</th>
<th>Protein/Organism/Length</th>
<th>NOV83a Residues/ Match Residues</th>
<th>Identities/ Similarities for the Matched Portion</th>
<th>Expect Value</th>
</tr>
<tr>
<td>Q93068</td>
<td>Ubiquitin-like protein SMT3C precursor (Ubiquitin-homology domain protein PIC1) (Ubiquitin-like protein UBL1) (Ubiquitin-related protein SUMO-1) (GAP modifying protein 1) (GMP1) (Sentrin) - Homo sapiens (Human), and, 101 aa.</td>
<td>1..101
1..101</td>
<td>89/101 (88%)
94/101 (92%)</td>
<td>6e-45</td>
</tr>
<tr>
<td>Q9MZD5</td>
<td>SENTRIN - Cervus nippon (Sika deer), 101 aa.</td>
<td>1..101
1..101</td>
<td>88/101 (87%)
93/101 (91%)</td>
<td>2e-44</td>
</tr>
<tr>
<td>O57686</td>
<td>SUMO-1 PROTEIN - Xenopus laevis (African clawed frog), 102 aa.</td>
<td>1..100
1..101</td>
<td>83/101 (82%)
90/101 (88%)</td>
<td>2e-39</td>
</tr>
<tr>
<td>Q9PT08</td>
<td>SMALL UBIQUITIN-RELATED PROTEIN 1 - Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri), 101 aa.</td>
<td>1..97
1..97</td>
<td>72/97 (74%)
84/97 (86%)</td>
<td>9e-35</td>
</tr>
<tr>
<td>Q9D466</td>
<td>4933411G06RIK PROTEIN - Mus musculus (Mouse), 117 aa.</td>
<td>1..97
1..96</td>
<td>68/97 (70%)
80/97 (82%)</td>
<td>8e-30</td>
</tr>
</table>

PFam analysis predicts that the NOV83a protein contains the domains shown in the Table 83F.

<table>
<tr><th colspan="4">Table 83F. Domain Analysis of NOV83a</th></tr>
<tr>
<th>Pfam Domain</th>
<th>NOV83a Match Region</th>
<th>Identities/ Similarities for the Matched Region</th>
<th>Expect Value</th>
</tr>
<tr>
<td>ubiquitin: domain 1 of 1</td>
<td>20..95</td>
<td>14/83 (17%)
66/83 (80%)</td>
<td>4.7e-18</td>
</tr>
</table>

Example 84.

The NOV84 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 84A.

Table 84A. NOV84 Sequence Analysis

	SEQ ID NO: 237	912 bp
NOV84a, CG59586-01 DNA Sequence	ACTCACTAATGGGCTCGAGCGGCTGCCTGTGTTTCAGCGGCTCGGGGAAATCCACCGT GGGCGCCCTGCTGGCATCTGAGCTGGGATGGAAATTCTATGATGCTGATGATTATCAC CCGGAGGAAAATCGAAGGAAGATGGGAAAAGGCATACCGCTCAATGACCAGGACCGGA TTCCATGGCTCTGTAACTTGCATGACATTTTACTAAGAGATGTAGCCTCGGGACAGCG TGTGGTTCTAGCCTGTTCAGCCCTGAAGAAAACGTACAGAGACATATTAACACAAGGA AAAGATGGTGTAGCTCTGAAGTGTGAGGAGTCGGGAAAGGAAGCAAAGCAGGCTGAGA TGCAGCTCCTGGTGGTCCATCTGAGCGGGTCGTTTGAGGTCATCTCTGGACGCTTACT CAAAAGAGAGGGACATTTTATGCCCCCTGAATTATTGCAGTCCCAGTTTGAGACTCTG GAGCCCCCAGCAGCTCCAGAAAACTTTATCCAAATAAGTGTGGACAAAAATGTTTCAG AGATAATTGCTACAATTATGGAAACCCTAAAAATGAAATGACAATGATTTTGTATCAG TGGTCCAAACAGAACTAAGCATAAATCATTGTGCCATCCCAAACCTCGTTCCAGCCGC CTTGCCCATACTAGATTCTAAATGTTTCTAAAGGCAAACCCCAATGTGTCAAGACAGA CTTGTTTAGGTGTAATTTTAGGAATTATGCTGGTTCATCAGGAAGCAGAGGGGGAGTT TTAAAAGTCAAGCTTAAATTGAAGTTTAAATTCATCTATAACCAAATCAAATGATCAG AGGAAATTCTGTAATCAATGCTGGAAATCGTTACATTGTTTAGAACATTCTTGCTCAT GCCTGTATTTGCACAAATAAATGAAACTTCGCTGTAAAAAAA
	ORF Start: ATG at 9	ORF Stop: TGA at 561
	SEQ ID NO: 238	184 aa	MW at 20352.2kD
NOV84a, CG59586-01 Protein Sequence	MGSSGCLCFSGSGKSTVGALLASELGWKFYDADDYHPEENRRKMGKGIPLNDQDRIPW LCNLHDILLRDVASGQRVVLACSALKKTYRDILTQGKDGVALKCEESGKEAKQAEMQL LVVHLSGSFEVISGRLLKREGHFMPPELLQSQFETLEPPAAPENFIQISVDKNVSEII ATIMETLKMK		

Further analysis of the NOV84a protein yielded the following properties shown in Table 84B.

Table 84B. Protein Sequence Properties NOV84a

PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.1000 probability located in plasma membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV84a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 84C.

Table 84C. Geneseq Results for NOV84a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV84a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG73989	Human colon cancer antigen protein SEQ ID NO:4753 - Homo sapiens, 193 aa. [WO200122920-A2, 05-APR-2001]	10..184 19..193	175/175 (100%) 175/175 (100%)	1e-97

342

AAB58998	Breast and ovarian cancer associated antigen protein sequence SEQ ID 706 - Homo sapiens, 193 aa. [WO200055173-A1, 21-SEP-2000]	10..184 19..193	175/175 (100%) 175/175 (100%)	1e-97
AAM89100	Human immune/haematopoietic antigen SEQ ID NO:16693 - Homo sapiens, 133 aa. [WO200157182-A2, 09-AUG-2001]	24..126 22..124	70/103 (67%) 77/103 (73%)	1e-34
AAG50675	Arabidopsis thaliana protein fragment SEQ ID NO: 64243 - Arabidopsis thaliana, 175 aa. [EP1033405-A2, 06-SEP-2000]	10..179 4..167	75/173 (43%) 102/173 (58%)	4e-28
AAG50674	Arabidopsis thaliana protein fragment SEQ ID NO: 64242 - Arabidopsis thaliana, 187 aa. [EP1033405-A2, 06-SEP-2000]	10..179 16..179	75/173 (43%) 102/173 (58%)	4e-28

In a BLAST search of public sequence databases, the NOV84a protein was found to have homology to the proteins shown in the BLASTP data in Table 84D.

Table 84D. Public BLASTP Results for NOV84a				
Protein Accession Number	Protein/Organism/Length	NOV84a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
BAB74785	GLUCONOKINASE - Anabaena sp. (strain PCC 7120), 160 aa.	10..183 9..160	72/174 (41%) 101/174 (57%)	1e-30
Q9RT56	THERMORESISTANT GLUCONOKINASE - Deinococcus radiodurans, 172 aa.	10..183 4..159	66/174 (37%) 101/174 (57%)	1e-29
CAC93415	PUTATIVE GLUCONOKINASE (EC 2.7.1.12) - Yersinia pestis, 167 aa.	10..174 12..159	68/166 (40%) 95/166 (56%)	2e-29
Q9CMM6	GLK - Pasteurella multocida, 172 aa.	10..182 15..169	68/174 (39%) 99/174 (56%)	2e-29
AAK86014	AGR_C_329P - Agrobacterium tumefaciens str. C58 (Cereon), 163 aa.	10..182 5..159	74/173 (42%) 98/173 (55%)	6e-29

PFam analysis predicts that the NOV84a protein contains the domains shown in the Table 84E.

Table 84E. Domain Analysis of NOV84a

Pfam Domain	NOV84a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SKI: domain 1 of 1	9..182	37/206 (18%) 114/206 (55%)	1.1

Example 85.

The NOV85 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 85A.

Table 85A. NOV85 Sequence Analysis

	SEQ ID NO: 239	4332 bp
NOV85a, CG59704-01 DNA Sequence	GGCGTATTAACGCGCGGGTGCACACCCCCACGGGCGCGCAATGAACAACTATGTGCTT AATGACGAGATCGGCCAGGGTGCCTTCAGCACTATTTACAAGGGCCGCTATCGCACCA CCACCGAGTTCTACGCGATTGCTTCCATCGACAAGAAGCGACGGGAGCGCGTCGTGAA CTGCGTTCAGCTGTTACGCTCCATGCACCACTCAAACGTCATAGAGTTCCACAACTGG TATGAGACCAACAATCACTTGTGGATCATTACGGAGTACTGCACCGGCGGAGACATGA GCACGATCCTCCGCTCGAACATTAATCTCACCACTCAGGCGGTCCAGGCGTTCGGCCG TGATGTGGCGATGGGCCTCATGTACATCCACAGTAAGGGTGTCGTGTATAACGACTTG CAGACTCGCAATCTGCTGATGGACTCCGCAGCAATGCTGCGCTTCCACGACTTTAGCT TGGCCTGTCTCTTCCAAGACGCGGCGACGCGGCCACTGGTGGGGACGCCACTGTACAT GGCCCCCGAGTTGTTCATGGCGGATCGCCCGCTGTACTCGATGGCATCAGACCTGTGG TCCTTCGGTTGTGTGCTGCACGAGCTGGCGACAGGCAAGCCGCCCTTTGCCGCATCCG ACCTCGAGACGCTGCTGGGCGACATACTGACGAGTCCGACGCCAGCGGTGCCTGGTGC GCCGGAGTCCTTTCAAACGCTCCTGTGCGGCCTGCTGGAAAAGGACCCGTTGAAGCGC TACGCGTGGGTCGATGTTGTCCGCAGCGAGTTCTGGGATGAGCCCTTGCCGCTGCCGA GCAACGGCTTTCCATCTCAGGTGGCGTGGGAGGACTACAAGCGTTCGCGTTCTGGACG CGGTGCGAGTCAGTATAATTGGACGGACTCCGATGTGCGTGTGGCAGTGGCTCACGCC GTGGGGGCAGCGAAATCAAACGCTTCTACGCACAACGTGGAGGAGAGGGAGCGAGCGG CTGCGACGTTGAACGTCGCGAAGGAGCTGGACTTCACTGCAAGCGCGGCGATGTTGCT GGAACGGTTACCGGAGCGGACACAGGAGCGTGCTGCGCACGCAACAGGCCATGTCGCG ACCGCGCACGGCAGCCTGGTGCACGGCTGCCCATCCACGGCCTCAGCGGCGGACCTCGC CAAGACGTTCAAGGACAAGGCGGCGCTGCTCAAGATTGTGGAAGAGGTCAAAACCGCT GTCGAGGGCTTCAAGCCGTGGGTGTCCTTCCACGTCGCTGCGCCACCCGGGCATGAGG GAGCGCCACTGGACCGGCTTGTCTCAGAAGCTCGGGATGAAGCTGGTGCCTGGCGACA CACTGATGCTTCTGGAGGACTGCGAACCGCTGCTAGCGCACCGCGACACCATTATCAG CTACTGCGAGGTGGCCGCGAAGGAGTCGCAGATCGAGATGACGCTCAAGGACATGCGT GCCAAGTGGGAGACCAAGTGCTTCATCATCGAGGCATACAAGGAGACAGGCACGTACA TCCTCAAGGACACCTCCGAGGTGGTGGAGCTCCTCGACGAGCACCTCAACGTCGTCCA GCAGCTGCAGTTCTCTCCATTCAAGGGCTACTTCGAGGAGTCCATCACGGACTGGGAG CGCTCCCTCAACCTCATCTCCGACATACTCGAACAATGGCTGGAGTGCCAGCGAGCGT GGCGTTATCTGGAGCCGATCCTCAACTCGGAGGACATCGCCATGCAGCTACCGCGACT GTCCACGCTGTTCGAGAAGGTGGACCGCACATGGAGACGTGTCATGGGCAACGCGCAC GCGCAGCCAAACGCACTCGAGTACTGCATTGGCACAAACAAGCTCTTGGACCACCTGC GCGAGGCGAACCGGCTCCTCGAAGTGCTGCAGCACTTGATGGCGCAGAAGGTCAACGT TGCCGCTGTTGGTCCGACTGGCACCGGCAAGTCCATCTCACTCGCGCGTCTCGTGCTT GGCGGCGGCATGCCGGCCAACTTTCTTGGCCTCAACTTCACCTTCTCGGCGCAGACAA AGTGCACAGTGTTGCAGAATTCACTGATGGCCAAGTTCGATAAGCGGCGCTCGCACGT CTACGGCGCCCCTGCCGGTAAGCACTTTCTCATCTTCATTGACGACGCGAACCTGCCG CAGCCAGAGAAGTACGGCGCGCAGCCCCCGGTGGAGCTTCTGCGGCAGATGCTCGCCC AAGGCGGCTTCTACAACTTTACAGGTGGCATCAAGTGGTCCTCCATCATCGACTGCTC GCTTGCGCTGGCGATGGGGCCGCCTGGCGGGGGCCGCAGCCGGGTTTCGAACCGCTTT ATGCGCTACTTCAATTACCTTGCCTTCCCCGAGATGTCGGACATGTCGAAGCGAACGA TCTTGCAGGCCATCCTCGTCGGCGGCCTCGCGCAGAGCGGCCTCGCTGACCGCCTCGC GAACGTCGCCTCCGCCGTGGTCGATAGCACGTTGCGGGTGTTTCGCAAGTGCACCCAG GTCTTTCTGCCGACCCCGGCGCACGTGCACTACTCCTTCAACATGCGGGATGTGATGC GTGTTTTTCCCCTCTTGTACACAGCAGACAAGTCGGTGCTGCAGTCGGAGGAATCCAT CGTGCGGCTGTGGATGCACGAGATGCAGCGCGTCTTCTACGATCGCCTCGTCGACGCG ACAGACAAGGGTCTGTTCATCGAGTACCTCAATGCCGAGCTGCCGTCCATGGGGGTGG ACAAGTCCTACAACGAGGTAGTGAAGGCTGACCGCCTCATCTTTGCCGACGTACTGAG CGACAAGGGCGTGTACGAGCAGATTACCGACATGAACGCCCTCACGACACGCATGAAT GAGCTGCTGGAGGCGTACAAATGACGAGAATGAAGTGAAGATGAACCTCGTGCTCTTCC TCGACGCCATCGAGCATGTCTGCCGTATCTCGCGCGTGCTGCGACTGCCGAACGGGCA CTGCCTCCTCCTCGGCGTTGGCGGGTCGGGACGCAAGTCACTCACGCGCCTGGCTTGT

344

	TCTCTGATTGCCGAGATGGAGGTGTTCACGATTGAGCTGTCGAAGAACTTCGGTGTCA
`	AGGAATGGCACGAGAGCCTCGCGAAGTTGCTGCTCGAGTGTGGCAAGGACGAGAAGAA
	GCGGACGTTTCTCTTCGCCGACACCCAGCTGGCGCATCCGACGTTTCTGGAGGATGTG
	GCGGGCCTGCTCACATCGGGTGATGTGCCGAACCTCTTTGAGGACCAAGATATCGAGC
	TCATCAACGACAAGTTTCGCGGCGTCTGCCTAAGCGAGAACCTGCCAACGACGAAGGT
	GTCGGTGTACGCGCGCTTTGTGAAGGAGGCGCGAGCCAACCTGCACCTTGTGCTCGCC
	TTCTCTCCCATCGGAGAGGCGTTTCGCAGCCGCCTGCGTATGTTCCCATCGCTCATTG
	CGTGCTGCACAATCGACTGGTTTGCTGAGTGGCCATCCGAGGCGCTACTGTCGGTAGC
	CGCAGTGCAGCTGAACGCCGGCGACGTTACTGACGTCATGGGGGCGGCAAGCCATGCC
	GACTTGCCGGGCTGCTTCCAGGCAGTGCACCGCGCGGCGGCGGAGGTGACGGAGCGCT
	TCTTCACGGAAACGCGTCGTCGCTCGTACGTGACGCCGACGTCCTATCTGTCGCTCCT
	CTCCAACTTCAAAGTGATGGCGGCGGCAAAACGCCGCTTCGTTCGCGAGCAGCGCGGC
	CGCCTCGAGAAGGGGCTGGAGAAGCTGCGGCACACCGAGGTGCAAGTGGCGGAGCTGG
	AGGCCCAGCTCAAGGCGCAGCAGCCGGTTCTGGTGCAGAAAAAGGCAGAGATTCAGTC
	GATGATGGAGCGGCTGACGGTGGACCGAAAGGAGGCGGCGGTGAAGGAGGCGGACGCG
	CGCAGGGAGGCCCAGCTTCCCGGTGGCCGTGCTGCTGCATACGGCGGTGAAGATGACGAAT
	GAGCCGCCGATGGGGCTGCGGGCGAACGTGATGCGCTCCTACTACGGCTTCACTCCCG
	AGGACCTCGAGCAGGAGGAGAAGCCCGCCGAGTTCAAAAAGATGTTGATGGCATCCGC
	ATGCCTGGTCCCATACCCGAGCACTGAAGAGCAGGGTCTCTGGAGCCTGGCATCGTGG
	GGTGGCCCTCAGCTTCCCCACTCACTGTGGGAAGTTTCCTTAGTGTCTCTGAGCCTGT
	TTCCTCATCCGTTGCCTGAGGATAAACCTGCTTCAGGATTGTTGGTGAAAAGACTTCC
	CTCACCTAGCTTCTGTAACGCCACTGCATGCCACCACTGCTGAGTACTGTTTGTTTGC
	TAGGTTGGTGTCATTCTCATTTTACCAGAAAGTGAAGCTC

	ORF Start: ATG at 41	ORF Stop: TGA at 3944	
	SEQ ID NO: 240	1301 aa	MW at 146115.7kD
NOV85a, CG59704-01 Protein Sequence	MNNYVLNDEIGQGAFSTIYKGRYRTTTEFYAIASIDKKRRERVVNCVQLLRSMHHSNV IEFHNWYETNNHLWIITEYCTGGDMSTILRSNINLTTQAVQAFGRDVAMGLMYIHSKG VVYNDLQTRNLLMDSAAMLRFHDFSLACLFQDAATRPLVGTPLYMAPELFMADRPLYS MASDLWSFGCVLHELATGKPPFAASDLETLLGDILTSPTPAVPGAPESFQTLLCGLLE KDPLKRYAWVDVVRSEFWDEPLPLPSNGFPSQVAWEDYKRSRSGRGASQYNWTDSDVR VAVAHAVGAAKSNASTHNVEERERAAATLNVAKELDFTASAAMLLERLPERTQERAAH ATGHVATAHGSLVHGCPSTASAATSPRRSRTRRRCSRLWKRSKPLSRASSRGCPSTSL RHPGMRERHWTGLSQKLGMKLVPGDTLMLLEDCEPLLAHRDTIISYCEVAAKESQIEM TLKDMRAKWETKCFIIEAYKETGTYILKDTSEVVELLDEHLNVVQQLQFSPFKGYFEE SITDWERSLNLISDILEQWLECQRAWRYLEPILNSEDIAMQLPRLSTLFEKVDRTWRR VMGNAHAQPNALEYCIGTNKLLDHLREANRLLEVLQHLMAQKVNVAAVGPTGTGKSIS LARLVLGGGMPANFLGLNFTFSAQTKCTVLQNSLMAKFDKRRSHVYGAPAGKHFLIFI DDANLPQPEKYGAQPPVELLRQMLAQGGFYNFTGGIKWSSIIDCSLALAMGPPGGGRS RVSNRFMRYFNYLAFPEMSDMSKRTILQAILVGGLAQSGLADRLANVASAVVDSTLRV FRKCTQVFLPTPAHVHYSFNMRDVMRVFPLLYTADKSVLQSEESIVRLWMHEMQRVFY DRLVDATDKGLFIEYLNAELPSMGVDKSYNEVVKADRLIFADVLSDKGVYEQITDMNA LTTRMNELLEAYNDENEVKMNLVLFLDAIEHVCRISRVLRLPNGHCLLLGVGGSGRKS LTRLACSLIAEMEVFTIELSKNFGVKEWHESLAKLLLECGKDEKKRTFLFADTQLAHP TFLEDVAGLLTSGDVPNLFEDQDIELINDKFRGVCLSENLPTTKVSVYARFVKEARAN LHLVLAFSPIGEAFRSRLRMFPSLIACCTIDWFAEWPSEALLSVAAVQLNAGDVTDVM GAASHADLPGCFQAVHRAAAEVTERFFTETRRRSYVTPTSYLSLLSNFKVMAAAKRRF VREQRGRLEKGLEKLRHTEVQVAELEAQLKAQQPVLVQKKAEIQSMMERLTVDRKEAA VKEADARREAQLPGGRAAYGGEDDE

Further analysis of the NOV85a protein yielded the following properties shown in Table 85B.

Table 85B. Protein Sequence Properties NOV85a	
PSort analysis:	0.8800 probability located in nucleus; 0.3562 probability located in microbody (peroxisome); 0.1671 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV85a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 85C.

Table 85C. Geneseq Results for NOV85a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV85a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM79863	Human protein SEQ ID NO 3509 - Homo sapiens, 2127 aa. [WO200157190-A2, 09-AUG-2001]	602..1287 168..847	218/692 (31%) 347/692 (49%)	1e-89
AAM79862	Human protein SEQ ID NO 3508 - Homo sapiens, 2127 aa. [WO200157190-A2, 09-AUG-2001]	602..1287 168..847	218/692 (31%) 347/692 (49%)	1e-89
AAM78879	Human protein SEQ ID NO 1541 - Homo sapiens, 2143 aa. [WO200157190-A2, 09-AUG-2001]	602..1287 108..787	218/692 (31%) 347/692 (49%)	1e-89
AAM78878	Human protein SEQ ID NO 1540 - Homo sapiens, 2067 aa. [WO200157190-A2, 09-AUG-2001]	602..1287 108..787	218/692 (31%) 347/692 (49%)	1e-89
AAM80293	Human protein SEQ ID NO 3945 - Homo sapiens, 1774 aa. [WO200157190-A2, 09-AUG-2001]	910..1293 33..405	153/393 (38%) 227/393 (56%)	5e-70

5   In a BLAST search of public sequence databases, the NOV85a protein was found to have homology to the proteins shown in the BLASTP data in Table 85D.

Table 85D. Public BLASTP Results for NOV85a				
Protein Accession Number	**Protein/Organism/Length**	**NOV85a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAL37427	CILIARY DYNEIN HEAVY CHAIN 7 - Homo sapiens (Human), 4024 aa.	628..1293 1975..2655	271/692 (39%) 395/692 (56%)	e-132
Q27812	DYNEIN HEAVY CHAIN ISOTYPE 7B (EC 3.6.1.3) - Tripneustes gratilla (Hawaian sea urchin), 1314 aa (fragment).	601..1247 654..1310	264/667 (39%) 389/667 (57%)	e-127
Q9MBF8	1 BETA DYNEIN HEAVY CHAIN - Chlamydomonas reinhardtii, 4513 aa.	611..1293 2486..3159	257/693 (37%) 377/693 (54%)	e-117
Q9VJC6	DHC36C PROTEIN - Drosophila melanogaster (Fruit fly), 4010 aa.	596..1275 1913..2604	249/699 (35%) 383/699 (54%)	e-116
Q9VWZ3	DHC16F PROTEIN - Drosophila melanogaster (Fruit fly), 4081 aa.	618..1301 2022..2709	248/704 (35%) 380/704 (53%)	e-108

PFam analysis predicts that the NOV85a protein contains the domains shown in the Table 85E.

Table 85E. Domain Analysis of NOV85a			
Pfam Domain	**NOV85a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
pkinase: domain 1 of 1	4..250	80/286 (28%) 190/286 (66%)	6.8e-62
DEAD: domain 1 of 1	613..637	7/25 (28%) 22/25 (88%)	0.83
dNK: domain 1 of 1	865..1020	32/179 (18%) 101/179 (56%)	6.8

Example 86.

5   The NOV86 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 86A.

Table 86A. NOV86 Sequence Analysis

	SEQ ID.NO: 241	1420 bp
NOV86a, CG59628-01 DNA Sequence	GTCCAGCTTTAGCTCTCTGCTCGCCGCCGCCGCTGTCGCCGCCACCTCCTCTGATCTA CGAAAGTCATGTTACCCAACACCGGGAGGCTGGCAGGATGTACAGTTTTTATCACAGG TGCAAGCCGTGGCATTGGCAAAGCTATTGCATTGAAAGCAGCAAAGGATGGAGCAAAT ATTGTTATTGCTGCAAAGACCGCCCAGCCACATCCAAAACTTCTAGGCACAATCTATA CTGCTGCTGAAGAAATTGAAGCAGTTGGAGGAAAGGCCTTGCCATGTATTGTTGATGT GAGAGATGAACAGCAGATCAGTGCTGCAGTGGAGAAAGCCATCAAGAAATTTGGAGGA ATTGATATTCTGGTAAATAATGCCAGTGCCATTAGTTTGACCAATACATTGGACACAC CTACCAAGAGATTGGATCTGATGATGAACGTGAACACCAGAGGCACCTACCTTGCATC TAAAGCATGTATTCCTTATTTGAAAAAGAGCAAAGTTGCTCATATCCTCAATATCAGT CCACCACTGAACCTAAATCCAGTTTGGTTCAAACAGCACTGTGCTTATACCATTGCTA AGTATGGTATGTCTATGTATGTGCTTGGAATGGCAGAAGAATTTAAAGGTGAAATTGC AGTCAATGCATTATGGCCTAAAACAGCCATACACTGCTGCTATGGATATGCTGGGA GGACCTGGTATCGAAAGCCAGTGTAGAAAAGTTGATATCATTGCAGATGCAGCATATT CCATTTTCCAAAAGCCAAAAAGTTTTACTGGCAACTTTGTCATTGATGAAAATATCTT AAAAGAAGAAGGAATAGAAAATTTTGACGTTTATGCAATTAAACCAGGTCATCCTTTG CAACCAGATTTCTTCTTAGATGAATACCCAGAAGCAGTTAGCAAGAAAGTGGAATCAA CTGGTGCTGTTCCAGAATTCAAAGAAGAGAAACTGCAGCTGCAACCAAAACCACGTTC TGGAGCTGTGGAAGAAACATTTAGAATTGTTAAGGACTCTCTCAGTGATGATGTTGTT AAAGCCACTCAAGCAATCTATCTGTTTGAACTCTCCGGTGAAGATGGTGGCACGTGGT TTCTTGATCTGAAAAGCAAGGGTGGAATGTCGGATATGGAGAGCCTTCTGATCAGGC AGATGTGGTGATGAGTATGACTACTGATGACTTTGTAAAAATGTTTTCAGGTAAACTA AAACCAACAATGGCATTCATGTCAGGGAAATTGAAGATTAAAGGTAACATGGCCCTAG CAATCAAATTGGAGAAGCTAATGAATCAGATGAATGCCAGACTGTGAAGGAAAATATA AAAAAAAAGTCGACTGCTATGCTCAAAAAGTAAAAAAAGCTCAACAGTTAAAATCTAA TGTTTGTTTTCTTTCCTGTTATATTATA
	ORF Start: ATG at 67	ORF Stop: TGA at 1321
	SEQ ID NO: 242	418 aa	MW at 45394.2kD
NOV86a, CG59628-01 Protein Sequence	MLPNTGRLAGCTVFITGASRGIGKAIALKAAKDGANIVIAAKTAQPHPKLLGTIYTAA EEIEAVGGKALPCIVDVRDEQQISAAVEKAIKKFGGIDILVNNASAISLTNTLDTPTK RLDLMMNVNTRGTYLASKACIPYLKKSKVAHILNISPPLNLNPVWFKQHCAYTIAKYG MSMYVLGMAEEFKGEIAVNALWPKTAIHTAAMDMLGGPGIESQCRKVDIIADAAYSIF QKPKSFTGNFVIDENILKEEGIENFDVYAIKPGHPLQPDFFLDEYPEAVSKKVESTGA VPEFKEEKLQLQPKPRSGAVEETFRIVKDSLSDDVVKATQAIYLFELSGEDGGTWFLD LKSKGGNVGYGEPSDQADVVMSMTTDDFVKMFSGKLKPTMAFMSGKLKIKGNMALAIK LEKLMNQMNARL

Further analysis of the NOV86a protein yielded the following properties shown in Table 86B.

Table 86B. Protein Sequence Properties NOV86a

PSort analysis:	0.5500 probability located in endoplasmic reticulum (membrane); 0.5000 probability located in microbody (peroxisome); 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV86a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 86C.

Table 86C. Geneseq Results for NOV86a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV86a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG81260	Human AFP protein sequence SEQ ID NO:38 - Homo sapiens, 418 aa. [WO200129221-A2, 26-APR-2001]	1..418 1..418	418/418 (100%) 418/418 (100%)	0.0
AAB84367	Amino acid sequence of human alcohol dehydrogenase 21612 - Homo sapiens, 418 aa. [WO200144446-A2, 21-JUN-2001]	1..418 1..418	418/418 (100%) 418/418 (100%)	0.0
AAG81258	Human AFP protein sequence SEQ ID NO:34 - Homo sapiens, 383 aa. [WO200129221-A2, 26-APR-2001]	1..382 1..382	382/382 (100%) 382/382 (100%)	0.0
ABB10251	Human cDNA SEQ ID NO: 559 - Homo sapiens, 278 aa. [WO200154474-A2, 02-AUG-2001]	141..418 1..278	271/278 (97%) 274/278 (98%)	e-156
AAU23020	Novel human enzyme polypeptide #106 - Homo sapiens, 278 aa. [WO200155301-A2, 02-AUG-2001]	141..418 1..278	271/278 (97%) 274/278 (98%)	e-156

In a BLAST search of public sequence databases, the NOV86a protein was found to have homology to the proteins shown in the BLASTP data in Table 86D.

Table 86D. Public BLASTP Results for NOV86a				
Protein Accession Number	Protein/Organism/Length	NOV86a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC38510	SEQUENCE 37 FROM PATENT WO0129221 - Homo sapiens (Human), 418 aa.	1..418 1..418	418/418 (100%) 418/418 (100%)	0.0
CAC38508	SEQUENCE 33 FROM PATENT WO0129221 - Homo sapiens (Human), 383 aa.	1..382 1..382	382/382 (100%) 382/382 (100%)	0.0

Q99LV2	HYPOTHETICAL 54.9 KDA PROTEIN - Mus musculus (Mouse), 496 aa.	1..418 1..496	355/496 (71%) 390/496 (78%)	0.0
Q9BT58	SIMILAR TO RIKEN CDNA 2610207I16 GENE - Homo sapiens (Human), 345 aa.	163..418 90..345	253/256 (98%) 254/256 (98%)	e-143
Q9VB10	CG5590 PROTEIN (GH01709P) - Drosophila melanogaster (Fruit fly), 412 aa.	4..418 3..412	238/422 (56%) 300/422 (70%)	e-128

PFam analysis predicts that the NOV86a protein contains the domains shown in the Table 86E.

Table 86E. Domain Analysis of NOV86a			
Pfam Domain	**NOV86a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
beta-lactamase: domain 1 of 1	222..236	4/15 (27%) 14/15 (93%)	6.5
adh_short: domain 1 of 1	9..321	74/339 (22%) 211/339 (62%)	2.4e-29
SCP2: domain 1 of 1	306..415	41/114 (36%) 87/114 (76%)	1.5e-25

Example 87.

The NOV87 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 87A.

Table 87A. NOV87 Sequence Analysis		
	SEQ ID NO: 243	888 bp
NOV87a, CG59516-01 DNA Sequence	TTCAACAAGGGCCCCTCCTACAGGCTCTTGGCGGACGTCCAGAACAGGCTTCTGTTCA AATATGACTCCCAGAAGGAGGCAGAGCTCCGCAGCTGGATCAAGGGATTCACTGGCCT CTCCATCCGCCCCGACTTCCAGAAGGGCCTGAAGGACGGGATTATTTTATGCACACTC GTGAACAAACTGCAGCCGGGCTCAGTCCCCAAGATCAACGGCTTCCGTGTAGAACTGG CACCAGCTAGAAAACCTCTCCAACATCCTCAAGGCAATGGTCAGCTACGGCATGATCC CGTGGACCTATTTGAGGCCAACGACCTGTTTGAGAGTGGGAACAATATGCAGGTGCGG GTGTCTCTTCTCGCCCTGGCAGGGAAGGCCAAGACTAAGGGGCTGCAGAGCGGGGTGG ACATCCGTGACAAGTACTCAGAGAAGCAGAACTTCAACGACACCACCATGAAGGCCAG GCTGTGCGTCATCCGGCTGCAGATTACCAACAAATGTGCCAGCCAGTCAGGCATGACC GCATACGTCACGAGGAGGCATCTCTACGACCCCAAGAACCGCATCCTGCCCCCCATGG ACAACTCGACCATCAGCCTCCGGATGGGTACAAACAAGTGCGCCAGCCAGGTGGGCAT GACGGCTCCCGGGAACCAGTGGCACATCTATGACACCAAGTTGGGAATCGACAAGTGT GAGAACTCCTCCATGTCCCTGAAGATGGGCTACACGCAGGTCGCCAATCACAGCAGAC AGGTCTTTGGCCTAGGCCGGCAAATATATGAACCCAAGTACCAGCCGGGTGGCCCAGT GGCCCACGGGGCTCCCTCCGCCGGCAACTGCCCAGGGCCAGGGGAGGCCCCTTAGTAC CAGGAGGAGACCAGCTAC
	ORF Start: TTC at 1	ORF Stop: TAG at 865

	SEQ ID NO: 244	288 aa	MW at 31831.2kD
NOV87a, CG59516-01 Protein Sequence	FNKGPSYRLLADVQNRLLFKYDSQKEAELRSWIKGFTGLSIRPDFQKGLKDGIILCTL VNKLQPGSVPKINGFRVELAPARKPLQHPQGNGQLRHDPVDLFEANDLFESGNNMQVR VSLLALAGKAKTKGLQSGVDIRDKYSEKQNFNDTTMKARLCVIRLQITNKCASQSGMT AYVTRRHLYDPKNRILPPMDNSTISLRMGTNKCASQVGMTAPGNQWHIYDTKLGIDKC ENSSMSLKMGYTQVANHSRQVFGLGRQIYEPKYQPGGPVAHGAPSAGNCPGPGEAP		

	SEQ ID NO: 245	888 bp	
NOV87b, CG59516-02 DNA Sequence	TTCAACAAGGGCCCCTCCTACAGGCTCTTGGCGGACGTCCAGAACAGGCTTCTGTTCA AATATGACTCCCAGAAGGAGGCAGAGCTCCGCAGCTGGATCAAGGGATTCACTGGCCT CTCCATCCGCCCCGACTTCCAGAAGGGCCTGAAGGACGGGATTATTTTATGCACACTC GTGAACAAACTGCAGCCGGGCTCAGTCCCCAAGATCAACGGCTTCCGTGTAGAACTGG CACCAGCTAGAAAACCTCTCCAACATCCTCAAGGCAATGGTCAGCTACGGCATGATCC CGTGGACCTATTTGAGGCCAACGACCTGTTTGAGAGTGGGAACAATATGCAGGTGCGG GTGTCTCTTCTCGCCCTGGCAGGGAAGGCCAAGACTAAGGGGCTGCAGAGCGGGGTGG ACATCCGTGACAAGTACTCAGAGAAGCAGAACTTCAACGACACCACCATGAAGGCCAG GCTGTGCGTCATCCGGCTGCAGATTACCAACAAATGTGCCAGCCAGTCAGGCATGACC GCATACGTCACGAGGAGGCATCTCTACGACCCCAAGAACCGCATCCTGCCCCCCATGG ACAACTCGACCATCAGCCTCCGGATGGGTACAAACAAGTGCGCCAGCCAGGTGGGCAT GACGGCTCCCGGGAACCAGTGGCACATCTATGACACCAAGTTGGGAATCGACAAGTGT GAGAACTCCTCCATGTCCCTGAAGATGGGCTACACGCAGGTCGCCAATCACAGCAGAC AGGTCTTTGGCCTAGGCCGGCAAATATAT̃G̃C̃CCCAAGTACCAGCCGGGTGGCCCAGT GGCCCACGGGGCTCCCTCCGCCGGCAACTG̃C̃CAGGGCCAGGGGAGGCCCCTTAGTAC CAGGAGGAGACCAGCTAC		

	ORF Start: TTC at 1	ORF Stop: TAG at 865	

	SEQ ID NO: 246	288 aa	MW at 31831.2kD
NOV87b, CG59516-02 Protein Sequence	FNKGPSYRLLADVQNRLLFKYDSQKEAELRSWIKGFTGLSIRPDFQKGLKDGIILCTL VNKLQPGSVPKINGFRVELAPARKPLQHPQGNGQLRHDPVDLFEANDLFESGNNMQVR VSLLALAGKAKTKGLQSGVDIRDKYSEKQNFNDTTMKARLCVIRLQITNKCASQSGMT AYVTRRHLYDPKNRILPPMDNSTISLRMGTNKCASQVGMTAPGNQWHIYDTKLGIDKC ENSSMSLKMGYTQVANHSRQVFGLGRQIYEPKYQPGGPVAHGAPSAGNCPGPGEAP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 87B.

Table 87B. Comparison of NOV87a against NOV87b.		
Protein Sequence	NOV87a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV87b	1..288 1..288	288/288 (100%) 288/288 (100%)

Further analysis of the NOV87a protein yielded the following properties shown in Table 87C.

Table 87C. Protein Sequence Properties NOV87a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.2110 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV87a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 87D.

Table 87D. Geneseq Results for NOV87a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV87a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR94888	Carponin - Homo sapiens, 297 aa. [JP08073380-A, 19-MAR-1996]	1..265 6..272	136/269 (50%) 176/269 (64%)	7e-63
AAR72588	Carponin protein - Homo sapiens, 297 aa. [WO9509010-A, 06-APR-1995]	1..265 6..272	136/269 (50%) 176/269 (64%)	7e-63
AAB43807	Human cancer associated protein sequence SEQ ID NO:1252 - Homo sapiens, 163 aa. [WO200055350-A1, 21-SEP-2000]	164..273 4..116	67/113 (59%) 82/113 (72%)	6e-30
AAM73074	Human bone marrow expressed probe encoded protein SEQ ID NO: 33380 - Homo sapiens, 71 aa. [WO200157276-A2, 09-AUG-2001]	157..225 2..71	49/70 (70%) 55/70 (78%)	4e-21
AAM60434	Human brain expressed single exon probe encoded protein SEQ ID NO: 32539 - Homo sapiens, 71 aa. [WO200157275-A2, 09-AUG-2001]	157..225 2..71	49/70 (70%) 55/70 (78%)	4e-21

In a BLAST search of public sequence databases, the NOV87a protein was found to have homology to the proteins shown in the BLASTP data in Table 87E.

Table 87E. Public BLASTP Results for NOV87a				
Protein Accession Number	Protein/Organism/Length	NOV87a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q08094	Calponin H2, smooth muscle - Sus scrofa (Pig), 296 aa (fragment).	1..287 6..296	219/291 (75%) 237/291 (81%)	e-116
Q99439	Calponin H2, smooth muscle (Neutral calponin) - Homo sapiens (Human), 309 aa.	1..288 6..297	218/292 (74%) 235/292 (79%)	e-115

Q08093	Calponin H2, smooth muscle - Mus musculus (Mouse), 305 aa.	1..288 6..293	214/291 (73%) 231/291 (78%)	e-112
O93547	CALPONIN H3 - Xenopus laevis (African clawed frog), 295 aa.	1..269 5..276	179/273 (65%) 208/273 (75%)	6e-91
Q922F8	UNKNOWN (PROTEIN FOR MGC:8135) - Mus musculus (Mouse), 242 aa.	59..288 1..230	166/233 (71%) 179/233 (76%)	8e-83

PFam analysis predicts that the NOV87a protein contains the domains shown in the Table 87F.

Table 87F. Domain Analysis of NOV87a			
Pfam Domain	NOV87a Match Region	Identities/ Similarities for the Matched Region	Expect Value
CH: domain 1 of 1	23..123	27/124 (22%) 65/124 (52%)	0.068
calponin: domain 1 of 2	159..183	17/26 (65%) 21/26 (81%)	3.8e-07
calponin: domain 2 of 2	198..223	15/26 (58%) 19/26 (73%)	3e-08

Example 88.

The NOV88 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 88A.

Table 88A. NOV88 Sequence Analysis		
	SEQ ID NO: 247	2213 bp
NOV88a, CG59671-02 DNA Sequence	CGTGAGGCACCCACTCTGGGAGCACAGAGAGCTCAGGTAGCCTGCCTAGATGGCGGCG CGCACCCTGGGCCGCGGCGTCGGGAGGCTGCTGGGCAGCCTGCGAGGGCTCTCGGGGC AGCCCGCGCGGCCGCCGTGCGGGGTGAGCGCGCCGCGCAGGGCGGCCTCGGGACCCTC GGGCAGCGCTCCCGCAGTTGCAGCAGCAGCAGCACAGCCAGGCTCGTATCCCGCGCTG AGTGCACAGGCAGCCCGGGAGCCGGCCGCCTTCTGGGGGCCTCTGGCGCGGGACACTC TCGTGTGGGACACCCCCTACCACACCGTCTGGGACTGCGACTTCAGCACTGGCAAGAT CGGCTGGTTCCTGGGAGGCCAGTTAAATGTCTCTGTCAACTGCTTGGACCAGCATGTT CGGAAGTCCCCCGAGAGCGTTGCTTTGATCTGGGAGCGCGATGAGCCTGGAACGGAAG TGAGGATCACCTACAGGGAACTACTGGAGACCACGTGCCGCCTGGCCAACACGCTGAA GAGGCATGGAGTCCACCGTGGGGACCGTGTTGCCATCTACATGCCCGTGTCCCATTG GCTGTGGCAGCAATGCTGGCCTGTGCCAGGATCGGAGCTGTCCACACAGTCATCTTTG CTGGCTTCAGTGCAGAGTCCTTGGCTGGGAGGATCAATGATGCCAAGTGCAAGGTGGT TATCACCTTCAACCAAGGACTCCGGGGTGGGCGCGTGGTGGAGCTGAAGAAAATAGTG GATGAGGCTGTGAAGCACTGCCCCACCGTGCAGCATGTCCTGGTGGCTCACAGGACAG ACAACAAGGTCCACATGGGGGATCTGGACGTCCCGCTGGAGCAGGAAATGGCCAAGGA GGACCCTGTTTGCGCCCCAGAGAGCATGGGCAGTGAGGACATGCTCTTCATGCTGTAC ACCTCAGGGAGCACCGGAATGCCCAAGGGCATCGTCCATACCCAGGCAGGCTACCTGC TCTATGCCGCCCTGACTCACAAGCTTGTGTTTGACCACCAGCCAGGTGACATCTTTGG CTGTGTGGCCGACATCGGTTGGATTACAGGACACAGCTACGTGGTGTATGGGCCTCTC TGCAATGGTGCCACCAGCGTCCTTTTTGAGAGCACCCCAGTTTATCCCAATGCTGGTC GGTACTGGGAGACAGTAGAGAGGTTGAAGATCAATCAGTTCTATGGCGCCCCAACGGC TGTCCGGCTGTTGCTGAAATACGGTGATGCCTGGGTGAAGAAGTATGATCGCTCCTCC

353

	CTGCGGACCCTGGGGTCAGTGGGAGAGCCCATCAACTGTGAGGCCTGGGAGTGGCTTC ACAGGGTGGTGGGGGACAGCAGGTGCACGCTGGTGGACACCTGGTGGCAGACAGAAAC AGGTGGCATCTGCATCGCACCACGGCCCTCGGAAGAAGGGGCGGAAATCCTCCCTGCC ATGGCGATGAGGCCCTTCTTTGGCATCGTCCCCGTCCTCATGGATGAGAAGGGCAGCG TCGTGGAGGGCAGCAACGTCTCCGGGGCCCTGTGCATCTCCCAGGCCTGGCCGGGCAT GGCCAGGACCATCTATGGCGACCACCAGCGATTTGTGGACGCCTACTTCAAGGCCTAC CCAGGCTATTACTTCACTGGAGACGGGGCTTACCGAACTGAGGGCGGCTATTACCAGA TCACAGGGCGGATGGATGATGTCATCAACATCAGTGGCCACCGGCTGGGGACCGCAGA GATTGAGGACGCCATCGCCGACCACCCTGCAGTACCAGAAAGTGCTGTCATTGGCTAC CCCCACGACATCAAAGGAGAAGCTGCCTTTGCCTTCATTGTGGTGAAAGATAGTGCGG GTGACTCAGATGTGGTGGTGCAGGAGCTCAAGTCCATGGTGGCCACCAAGATCGCCAA ATATGCTGTGCCTGATGAGATCCTGGTGGTGAAACGTCTTCCAAAAACCAGGTCTGGG AAGGTCATGCGGCGGCTCCTGAGGAAGATCATCACTAGTGAGGCCCAGGAGCTGGGAG ACACTACCACCTTGGAGGACCCCAGCATCATCGCAGAGATCCTGAGTGTCTACCAGAA GTGCAAGGACAAGCAGGCTGCTGCTAAGTGAGCTGGCACCTTGTGGGGCTCTTGGGAT GGGCGGGCACCCAAGCCCTGGCTTGTCCTTCCCAGAAGGTACCCCTGAGGTTGGCGTC TTCCTACGT
	ORF Start: ATG at 50	ORF Stop: TGA at 2117	
	SEQ ID NO: 248	689 aa	MW at 74855.9kD
NOV88a, CG59671-02 Protein Sequence	MAARTLGRGVGRLLGSLRGLSGQPARPPCGVSAPRRAASGPSGSAPAVAAAAAQPGSY PALSAQAAREPAAFWGPLARDTLVWDTPYHTVWDCDFSTGKIGWFLGGQLNVSVNCLD QHVRKSPESVALIWERDEPGTEVRITYRELLETTCRLANTLKRHGVHRGDRVAIYMPV SPLAVAAMLACARIGAVHTVIFAGFSAESLAGRINDAKCKVVITFNQGLRGGRVVELK KIVDEAVKHCPTVQHVLVAHRTDNKVHMGDLDVPLEQEMAKEDPVCAPESMGSEDMLF MLYTSGSTGMPKGIVHTQAGYLLYAALTHKLVFDHQPGDIFGCVADIGWITGHSYVVY GPLCNGATSVLFESTPVYPNAGRYWETVERLKINQFYGAPTAVRLLLKYGDAWVKKYD RSSLRTLGSVGEPINCEAWEWLHRVVGDSRCTLVDTWWQTETGGICIAPRPSEEGAEI LPAMAMRPFFGIVPVLMDEKGSVVEGSNVSGALCISQAWPGMARTIYGDHQRFVDAYF KAYPGYYFTGDGAYRTEGGYYQITGRMDDVINISGHRLGTAEIEDAIADHPAVPESAV IGYPHDIKGEAAFAFIVVKDSAGDSDVVVQELKSMVATKIAKYAVPDEILVVKRLPKT RSGKVMRRLLRKIITSEAQELGDTTTLEDPSIIAEILSVYQKCKDKQAAAK

Further analysis of the NOV88a protein yielded the following properties shown in Table 88B.

Table 88B. Protein Sequence Properties NOV88a	
PSort analysis:	0.6500 probability located in plasma membrane; 0.6000 probability located in nucleus; 0.4340 probability located in mitochondrial inner membrane; 0.3000 probability located in Golgi body
SignalP analysis:	Likely cleavage site between residues 23 and 24

A search of the NOV88a protein against the Geneseq database, a proprietary

5 database that contains sequences published in patents and patent publication, yielded

several homologous proteins shown in Table 88C.

Table 88C. Geneseq Results for NOV88a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV88a, Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU23058	Novel human enzyme polypeptide #144 - Homo sapiens, 664 aa. [WO200155301-A2, 02-AUG-2001]	26..689 1..664	663/664 (99%) 663/664 (99%)	0.0
AAB34712	Human secreted protein encoded by DNA clone vo9 1 - Homo sapiens, 518 aa. [WO200055375-A1, 21-SEP-2000]	172..689 1..518	518/518 (100%) 518/518 (100%)	0.0
AAU23050	Novel human enzyme polypeptide #136 - Homo sapiens, 479 aa. [WO200155301-A2, 02-AUG-2001]	224..689 18..479	459/466 (98%) 461/466 (98%)	0.0
ABB12253	Human acetate-coA ligase homologue, SEQ ID NO:2623 - Homo sapiens, 446 aa. [WO200157188-A2, 09-AUG-2001]	1..446 1..446	446/446 (100%) 446/446 (100%)	0.0
AAR23968	facA gene product - Penicillium chrysogenum, 669 aa. [WO9207079-A, 30-APR-1992]	58..684 45..669	305/629 (48%) 407/629 (64%)	e-175

In a BLAST search of public sequence databases, the NOV88a protein was found to have homology to the proteins shown in the BLASTP data in Table 88D.

Table 88D. Public BLASTP Results for NOV88a				
Protein Accession Number	Protein/Organism/Length	NOV88a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99NB1	ACETYL-COA SYNTHETASE 2 - Mus musculus (Mouse), 682 aa.	1..687 1..680	599/687 (87%) 638/687 (92%)	0.0
Q9BEA3	ACETYL-COA SYNTHETASE 2 - Bos taurus (Bovine), 675 aa.	1..689 1..675	575/689 (83%) 625/689 (90%)	0.0
Q9NUB1	DJ568C11.3 (NOVEL AMP-BINDING ENZYME SIMILAR TO ACETYL-COENZYME A	212..689 1..478	478/478 (100%) 478/478 (100%)	0.0

	LIGASE)) - Homo sapiens (Human), 478 aa (fragment).			
Q96JI1	KIAA1846 PROTEIN - Homo sapiens (Human), 354 aa (fragment).	336..689 1..354	354/354 (100%) 354/354 (100%)	0.0
Q9HV66	ACETYL-COENZYME A SYNTHETASE - Pseudomonas aeruginosa, 645 aa.	58..675 24..639	326/619 (52%) 433/619 (69%)	0.0

PFam analysis predicts that the NOV88a protein contains the domains shown in the Table 88E.

Table 88E. Domain Analysis of NOV88a			
Pfam Domain	NOV88a Match Region	Identities/ Similarities for the Matched Region	Expect Value
AMP-binding: domain 1 of 1	142..580	121/441 (27%) 341/441 (77%)	7.1e-117

Example 89.

The NOV89 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 89A.

Table 89A. NOV89 Sequence Analysis		
	SEQ ID NO: 249	1268 bp
NOV89a, CG56870-01 DNA Sequence	ACTTCTTTCTTTTCTGTTTCAGAGTTACTGATTTATTCTTGAGATTCCTCTACTCTCG TTATCTGACCTCATGGATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTC TAAATGATAAGAATGGTACAAGAAACTTCCAGGACTTTGACTGTCAGGAACATGATAT AGAAACAACTCATGGTGTGGTCCACGTCACTATAAGAGGCTTACCCAAAGGAAACAGA CCAGTTATACTAACATATCATGACATTGGCCTCAACCGTAAATCCTGTTTCAATGCAT TCTTTAACTTTGAGGATATGCAAGAGATCACCCAGCACTTTGCTGTCTGTCATGTGGA TGCCCCAGGCCAGCAGGAAGGTGCACCCTCTTTCCCAACAGGGTATCAGTACCCCACA ATGGATGAGCTGGCTGAAATGCTGCCTCCTGTTCTTACCCACCTAAGCCTGAAAAGCA TCATTGGAATTGGAGTTGGAGCTGGAGCTTACATCCTCAGCAGATTTGCACTCAACCA TCCAGAGCTTGTGGAAGGCCTTGTGCTCATTAATGTTGACCCTTGCGCTAAAGGCTGG ATTGACTGGGCAGCTTCCAAACTCTCTGGCCTGACAACCAATGTTGTGGACATTATTT TGGCTCATCACTTTGGGCAGGAAGAGTTACAGGCCAACCTGGACCTGATCCAAACCTA CAGAATGCATATTGCCCAAGACATCAACCAAGACAACCTGCAGCTCTTCTTGAATTCC TACAATGGGCGCAGAGACCTGGAGATCGAAAGACCCATACTGGGCCAAAATGATAACA AATCAAAAACATTAAAGTGTTCTACTTTACTGGTGGTAGGGGACAATTCGCCTGCAGT TGAGGCTGTGGTCGAATGCAATTCCCGCCTGAACCCTATAAATACAACTTTGCTAAAG ATGGCGGACTGTGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGCTCACCGAGGCCT TCAAGTACTTTTTGCAGGGAATGGGCTACGTCCCGTCTGCCAGCATGACTCGGCTCGC CCGATCACGAACCCACTCAACCTCGAGTAGCCTCGGCTCTGGAGAAAGTCCCTTCAGC CGGTCTGTCACCAGCAATCAGTCAGATGGAACTCAAGAATCCTGTGAGTCCCCTGATG TCCTGGACAGACACCAGACCATGGAGGTGTCCTGCTAAGCAGATGCTCCTCCCCTGGA CCATTGCAAGTCCATCCTTCAAATGACCACTCCATAATATAACATTTCAT
	ORF Start: ATG at 71	ORF Stop: TAA at 1196
	SEQ ID NO: 250	375 aa MW at 41413.3kD
NOV89a, CG56870-01 Protein Sequence	MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVIL TYHDIGLNRKSCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDEL AEMLPPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWA

356

	ASKLSGLTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGR RDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADC GGLPQVVQPGKLTEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVT SNQSDGTQESCESPDVLDRHQTMEVSC
	SEQ ID NO: 251 1175 bp
NOV89b, CG56870-02 DNA Sequence	<u>TCGTTATCTGACCTC</u>**ATG**GATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCAC TTCTAAATGATAAGAATGGTACAAGAAACTTCCAGGACTTTGACTGTCAGGAACATGA TATAGAAACAACTCATGGTGTGGTCCACGTCACTATAAGAGGCTTACCCAAAGGAAAC AGACCAGTTATACTAACATATCATGACATTGGCCTCAACCATAAATCCTGTTTCAATG CATTCTTTAACTTTGAGGATATGCAAGAGATCACCCAGCACTTTGCTGTCTGTCATGT GGATGCCCCAGGCCAGCAGGAAGGTGCACCCTCTTTCCCAACAGGGTATCAGTACCCC ACAATGGATGAGCTGGCTGAAGTGCTGCCTCCTGTTCTTACCCACCTAAGCCTGAAAA GCATCATTGGAATTGGAGTTGGAGCTGGAGCTTACATCCTCAGCAGATTTGCACTCAA CCATCCAGAGCTTGTGGAAGGCCTTGTGCTCATTAATGTTGACCCTTGCGCTAAAGGC TGGATTGACTGGGCAGCTTCCAAACTCTCTGGCCTGACAACCAATGTTGTGGACATTA TTTTGGCTCATCACTTTGGGCAGGAAGAGTTACAGGCCAACCTGGACCTGATCCAAAC CTACAGAATGCATATTGCCCAAGACATCAACCAAGCAACCTGCAGCTCTTCTTGAAT TCCTACAATGGACGCAGAGACCTGGAGATCGAAAGACCCATACTGGGCCAAAATGATA ACAAATCAAAAACATTAAAGTGTTCTACTTTACTGGTGGTAGGGGACAATTCGCCTGC AGTTGAGGCTGTGGTCGAATGCAATTCCCGCCTGAACCCTATAAATACAACTTTGCTA AAGATGGCGGACTGTGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGCTCACCGAGG CCTTCAAGTACTTTTTGCAGGGAATGGGCTACATACCATCTGCCAGCATGACTCGGCT CGCCCGATCACGAACCCACTCAACCTCGAGTAGCCTCGGCTCTGGAGAAAGTCCCTTC AGCCGGTCTGTCACCAGCAATCAGTCAGATGGAACTCAAGAATCCTGTGAGTCCCCTG ATGTCCTGGACAGACACCAGACCATGGAGGTGTCCTGC**TAA**<u>GCAGATGCTCCTCCCCT GGACCATTGCAAGTC</u>
	ORF Start: ATG at 16 **ORF Stop: TAA at 1141**
	SEQ ID NO: 252 375 aa MW at 41376.2kD
NOV89b, CG56870-02 Protein Sequence	MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVIL TYHDIGLNHKSCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDEL AEVLPPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWA ASKLSGLTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGR RDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADC GGLPQVVQPGKLTEAFKYFLQGMGYIPSASMTRLARSRTHSTSSSLGSGESPFSRSVT SNQSDGTQESCESPDVLDRHQTMEVSC
	SEQ ID NO: 253 1232 bp
NOV89c, CG56870-03 DNA Sequence	ACTTCTTTCTTTTCTGTTTCAGAGTTACTGATTTATTCTTGAGATTCCTCTACTCTCG <u>TTATCTGACCTC</u>**ATG**GATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTC TAAATGATAAGGAACATGATATAGAAACAACTCATGGTGTGGTCCACGTCACTATAAG AGGCTTACCCAAAGGAAACAGACCAGTTATACTAACATATCATGACATTGGCCTCAAC CATAAATCCTGTTTCAATGCATTCTTTAACTTTGAGGATATGCAAGAGATCACCCAGC ACTTTGCTGTCTGTCATGTGGATGCCCCAGGCCAGCAGGAAGGTGCACCCTCTTTCCC AACAGGGTATCAGTACCCCACAATGGATGAGCTGGCTGAAATGCTGCCTCCTGTTCTT ACCCACCTAAGCCTGAAAAGCATCATTGGAATTGGAGTTGGAGCTGGAGCTTACATCC TCAGCAGATTTGCACTCAACCATCCAGAGCTTGTGGAAGGCCTTGTGCTCATTAATGT TGACCCTTGCGCTAAAGGCTGGATTGACTGGGCAGCTTCCAAACTCTCTGGCCTGACA ACCAATGTTGTGGACATTATTTTGGCTCATCACTTTGGGCAGGAAGAGTTACAGGCCA ACCTGGACCTGATCCAAACCTACAGAATGCATATTGCCCAAGACATCAACCAAGACAA CCTGCAGCTCTTCTTGAATTCCTACAATGGGCGCAGAGACCTGGAGATCGAAAGACCC ATACTGGGCCAAAATGATAACAAATCAAAAACATTAAAGTGTTCTACTTTACTGGTGG TAGGGGACAATTCGCCTGCAGTTGAGGCTGTGGTCGAATGCAATTCCCGCCTGAACCC TATAAATACAACTTTGCTAAAGATGGCGGACTGTGGGGGACTGCCCCAGGTAGTTCAG CCTGGGAAGCTCACCGAGGCCTTCAAGTACTTTTTGCAGGGAATGGGCTACGTCCCGT CTGCCAGCATGACTCGGCTCGCCCGATCACGAACCCACTCAACCTCGAGTAGCCTCGG CTCTGGAGAAAGTCCCTTCAGCCGGTCTGTCACCAGCAATCAGTCAGATGGAACTCAA GAATCCTGTGAGTCCCCTGATGTCCTGGACAGACACCAGACCATGGAGGTGTCCTGCT **AAG**CAGATGCTCCTCCCCTGGACCATTGCAAGTCCATCCTTCAAATGACCACTCCATA ATATAACATTTCAT
	ORF Start: ATG at 71 **ORF Stop: TAA at 1160**
	SEQ ID NO: 254 363 aa MW at 39967.8kD
NOV89c, CG56870-03 Protein Sequence	MDELQDVQLTEIKPLLNDKEHDIETTHGVVHVTIRGLPKGNRPVILTYHDIGLNHKSC FNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEMLPPVLTHLS LKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSGLTTNVV DIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERPILGQ NDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADCGGLPQVVQPGKL TEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCE SPDVLDRHQTMEVSC
	SEQ ID NO: 255 1220 bp

NOV89d, CG56870-04 DNA Sequence	ACTTCTTTCTTTTCTGTTTCAGAGTTACTGATTTATTCTTGAGATTCCTCTACTCTCG TTATCTGACCTCATGGATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTC TAAATGATAAGAATGGTACAAGAAACTTCCAGGACTTTGACTGTCAGGAACATGATAT AGAAACAACTCATGGTGTGGTCCACGTCACTATAAGAGGCTTACCCAAAGGAAACAGA CCAGTTATACTAACATATCATGACATTGGCCTCAACCGTAAATCCTGTTTCAATGCAT TCTTTAACTTTGAGGATATGCAAGAGATCACCCAGCACTTTGCTGTCTGTCATGTGGA TGCCCCAGGCCAGCAGGAAGGTGCACCCTCTTTCCCAACAGGGTATCAGTACCCCACA ATGGATGAGCTGGCTGAAATGCTGCCTCCTGTTCTTACCCACCTAAGCCTGAAAAGCA TCATTGGAATTGGAGTTGGAGCTGGAGCTTACATCCTCAGCAGATTTGCACTCAACCA TCCAGAGCTTGTGGAAGGCCTTGTGCTCATTAATGTTGACCCTTGCGCTAAAGGCTGG ATTGACTGGGCAGCTTCCAAACTCTCTGGCCTGACAACCAATGTTGTGGACATTATTT TGGCTCATCACTTTGGGCAGGAAGAGTTACAGGCCAACCTGGACCTGATCCAAACCTA CAGAATGCATATTGCCCAAGACATCAACCAAGACAACCTGCAGCTCTTCTTGAATTCC TACAATGGACGCAGAGACCTGGAGATCGAAAGACCCATACTGGGCCAAAATGATAACA AATCAAAAACATTAAAGTGTTCTACTTTACTGGTGGTAGGGGACAATTCGCCTGCAGT TGAGGCTGTGATGGCGGACTGTGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGTTC ACCGAGGCCTTCAAGTACTTTTTGCAGGGAATGGGCTACACACCATCTGCCAGCATGA CTCGGCTCGCCCGATCACGAACCCACTCAACCTCGAGTAGCCTCGGCTCTGGAGAAAG TCCCTTCAGCCGGTCTGTCACCAGCAATCAGTCAGATGGAACTCAAGAATCCTGTGAG TCCCCTGATGTCCTGGACAGACACCAGACCATGGAGGTGTCCTGC**TAA**GCAGATGCTC CTCCCCTGGACCATTGCAAGTCCATCCTTCAAATGACCACTCCATAATATAACATTTC AT
	ORF Start: ATG at 71	ORF Stop: TAA at 1148
	SEQ ID NO: 256	359 aa	MW at 39652.2kD
NOV89d, CG56870-04 Protein Sequence	MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVIL TYHDIGLNRKSCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDEL AEMLPPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWA ASKLSGLTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGR RDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVMADCGGLPQVVQPGKFTEAF KYFLQGMGYTPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCESPDV LDRHQTMEVSC
	SEQ ID NO: 257	970 bp
NOV89e, CG56870-05 DNA Sequence	**ATG**GATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTCTAAATGATAAGA ATGGTACAAGAAACTTCCAGGACTTTGACTGTCAGTATCAGTACCCCACAATGGATGA GCTGGCTGAAATGCTGCCTCCTGTTCTTACCCACCTAAGCCTGAAAAGCATCATTGGA ATTGGAGTTGGAGCTGGAGCTTACATCCTCAGCAGATTTGCACTCAACCATCCAGAGC TTGTGGAAGGCCTTGTGCTCATTAATGTTGACCCTTGCGCTAAAGGCTGGATTGACTG GGCAGCTTCCAAACTCTCTGGCCTGACAACCAATGTTGTGGACATTATTTTGGCTCAT CACTTTGGGCAGGAAGAGTTACAGGCCAACCTGGACCTGATCCAAACCTACAGAATGC ATATTGCCCAAGACATCAACCAAGACAACCTGCAGCTCTTCTTGAATTCCTACAATGG GCGCAGAGACCTGGAGATCGAAAGACCCATACTGGGCCAAAATGATAACAAATCAAAA ACATTAAAGTGTTCTACTTTACTGGTGGTAGGGGACAATTCGCCTGCAGTTGAGGCTG TGGTCGAATGCAATTCCCGCCTGAACCCTATAAATACAACTTTGCTAAAGATGGCGGA CTGTGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGCTCACCGAGGCCTTCAAGTAC TTTTTGCAGGGAATGGGCTACGTCCCGTCTGCCAGCATGACTCGGCTCGCCCGATCAC GAACCCACTCAACCTCGAGTAGCCTCGGCTCTGGAGAAAGTCCCTTCAGCCGGTCTGT CACCAGCAATCAGTCAGATGGAACTCAAGAATCCTGTGAGTCCCCTGATGTCCTGGAC AGACACCAGACCATGGAGGTGTCCTGC**TAA**GCAGATGCTCCTCCCCTGGACCATTGCA AGTCCATCCTTCAAATGACCACTCCATAATATAACATTTCAT
	ORF Start: ATG at 1	ORF Stop: TAA at 898
	SEQ ID NO: 258	299 aa	MW at 32956.9kD
NOV89e, CG56870-05 Protein Sequence	MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQYQYPTMDELAEMLPPVLTHLSLKSIIG IGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSGLTTNVVDIILAH HFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERPILGQNDNKSK TLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADCGGLPQVVQPGKLTEAFKY FLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCESPDVLD RHQTMEVSC

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 89B.

358

Table 89B. Comparison of NOV89a against NOV89b through NOV89e.

Protein Sequence	NOV89a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV89b	1..375 1..375	336/375 (89%) 338/375 (89%)
NOV89c	1..375 1..363	326/375 (86%) 326/375 (86%)
NOV89d	1..375 1..359	321/375 (85%) 321/375 (85%)
NOV89e	104..375 28..299	233/272 (85%) 233/272 (85%)

Further analysis of the NOV89a protein yielded the following properties shown in Table 89C.

Table 89C. Protein Sequence Properties NOV89a

PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1685 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV89a protein against the Geneseq database, a proprietary
5   database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 89D.

Table 89D. Geneseq Results for NOV89a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV89a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM94019	Human stomach cancer expressed polypeptide SEQ ID NO 108 - Homo sapiens, 363 aa. [WO200109317-A1, 08-FEB-2001]	1..375 1..363	360/375 (96%) 361/375 (96%)	0.0
AAG64392	Human reducing agent and tunicamycin-responsive protein 40 - Homo sapiens, 363 aa. [WO200155375-A1, 02-AUG-2001]	1..375 1..363	360/375 (96%) 361/375 (96%)	0.0

AAB94494	Human protein sequence SEQ ID NO:15186 - Homo sapiens, 363 aa. [EP1074617-A2, 07-FEB-2001]	1..375 1..363	360/375 (96%) 361/375 (96%)	0.0
AAU31598	Novel human secreted protein #2089 - Homo sapiens, 395 aa. [WO200179449-A2, 25-OCT-2001]	68..374 1..307	282/323 (87%) 286/323 (88%)	e-154
AAB95462	Human protein sequence SEQ ID NO:17944 - Homo sapiens, 286 aa. [EP1074617-A2, 07-FEB-2001]	133..375 44..286	240/243 (98%) 242/243 (98%)	e-138

In a BLAST search of public sequence databases, the NOV89a protein was found to have homology to the proteins shown in the BLASTP data in Table 89E.

Table 89E. Public BLASTP Results for NOV89a				
Protein Accession Number	Protein/Organism/Length	NOV89a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UGV2	NDRG3 protein - Homo sapiens (Human), 375 aa.	1..375 1..375	373/375 (99%) 374/375 (99%)	0.0
Q96PL8	NDR1-RELATED DEVELOPMENT PROTEIN NDR3 - Homo sapiens (Human), 375 aa.	1..375 1..375	372/375 (99%) 373/375 (99%)	0.0
Q9QYF9	NDRG3 protein (Ndr3 protein) - Mus musculus (Mouse), 375 aa.	1..375 1..375	358/375 (95%) 368/375 (97%)	0.0
AAH18504	SIMILAR TO N-MYC DOWNSTREAM REGULATED 3 - Mus musculus (Mouse), 388 aa.	1..375 1..388	359/388 (92%) 368/388 (94%)	0.0
Q96SM2	CDNA FLJ14759 FIS, CLONE NT2RP3003290, MODERATELY SIMILAR TO MUS MUSCULUS NDR1 RELATED PROTEIN NDR3 - Homo sapiens (Human), 363 aa.	1..375 1..363	360/375 (96%) 361/375 (96%)	0.0

PFam analysis predicts that the NOV89a protein contains the domains shown in the Table 89F.

5

Table 89F. Domain Analysis of NOV89a			
Pfam Domain	NOV89a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Orn_Arg_deC_N: domain 1 of 1	62..89	7/33 (21%) 24/33 (73%)	1.9
abhydrolase: domain 1 of 1	87..310	48/239 (20%) 142/239 (59%)	0.0066
Ndr: domain 1 of 1	22..346	210/340 (62%) 311/340 (91%)	3.7e-211

Example 90.

The NOV90 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 90A.

Table 90A. NOV90 Sequence Analysis		
	SEQ ID NO: 259	632 bp
NOV90a, CG59764-01 DNA Sequence	GAAACTATAAAGGGTCCGAACCCTCTTTTAAAGGATCCCAATGCATTTCTTTGATCCC TCGCCGGTGCGACGGTACCACCATCCCAGCTGTGAGGCTGCCATCAACACCCACATCA GCCTGGAGCTCCACGCATCCTATGTGTACCTGTCCATGGCCTTCTACTTCGACCAGGA CGACGCGGCCCTGGAGCACTTTGACCGCTACTTCCTGCGCCAGTCGCAGGAGAAAAGG GAGCACGCCCAGGAGCTGATGAGCCTGCAGAACCTGCGCGGTGGCCGCATCTGCCTTC ATGACATCAGGAAGCCAGAGGGCCAAGGCTGGGAGAGCGGGCTCAAGGCCATGGAGTG CACCTTCCACCTGGAGAAGAACATCAACCAGAGCCTCCTGGAGCTGCACCAGCTGGCC AGGGAGAACGGCGACCCCCAGCTCTGCGACTTCCTGGAGAACGACTTCCTGAACCAGC AGGCCAAGACCATCAAAGAGCTGGGTGGCTACCTGAGCAACCTGCACAAGATGGGGGC CCCGGAAGCAGGCCTGGCAGAGTACCTCTTTAACAAGCTCACCCTGGGCCGCAGCGAA CCACTTCCTTGAACCAGCAGGCCAAGACCATCAAAGAGATTGGTGGCTACCT	
	ORF Start: ATG at 41	ORF Stop: TGA at 590
	SEQ ID NO: 260	183 aa	MW at 21159.6kD
NOV90a, CG59764-01 Protein Sequence	MHFFDPSPVRRYHHPSCEAAINTHISLELHASYVYLSMAFYFDQDDAALEHFDRYFLR QSQEKREHAQELMSLQNLRGGRICLHDIRKPEGQGWESGLKAMECTFHLEKNINQSLL ELHQLARENGDPQLCDFLENDFLNQQAKTIKELGGYLSNLHKMGAPEAGLAEYLFNKL TLGRSEPLP	

Further analysis of the NOV90a protein yielded the following properties shown in Table 90B.

Table 90B. Protein Sequence Properties NOV90a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.1400 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

361

A search of the NOV90a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 90C.

Table 90C. Geneseq Results for NOV90a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV90a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU07889	Polypeptide sequence for human hspG34a - Homo sapiens, 221 aa. [WO200166752-A2, 13-SEP-2001]	7..180 45..218	159/174 (91%) 164/174 (93%)	4e-91
AAU07890	Polypeptide sequence for human hspG34b - Homo sapiens, 183 aa. [WO200166752-A2, 13-SEP-2001]	6..177 6..177	125/172 (72%) 149/172 (85%)	6e-70
AAB90804	Human shear stress-response protein SEQ ID NO: 108 - Homo sapiens, 183 aa. [WO200125427-A1, 12-APR-2001]	7..180 7..180	114/174 (65%) 141/174 (80%)	6e-64
AAR71567	Human monocyte growth factor - Homo sapiens, 183 aa. [JP07031482-A, 03-FEB-1995]	7..180 7..180	114/174 (65%) 141/174 (80%)	6e-64
AAU27741	Mouse full-length polypeptide sequence #66 - Mus musculus, 182 aa. [WO200164834-A2, 07-SEP-2001]	6..180 6..180	112/175 (64%) 141/175 (80%)	5e-63

In a BLAST search of public sequence databases, the NOV90a protein was found to have homology to the proteins shown in the BLASTP data in Table 90D.

Table 90D. Public BLASTP Results for NOV90a				
Protein Accession Number	Protein/Organism/Length	NOV90a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BXU8	Ferritin heavy polypeptide-like 17 - Homo sapiens (Human), 183 aa.	6..177 6..177	125/172 (72%) 149/172 (85%)	2e-69
P29389	Ferritin heavy chain (Ferritin H subunit) - Cricetulus griseus (Chinese hamster), 185 aa.	6..180 10..184	115/175 (65%) 142/175 (80%)	6e-64

A26886	ferritin heavy chain - chicken, 180 aa.	6..180 5..179	112/175 (64%) 142/175 (81%)	1e-63
P08267	Ferritin heavy chain (Ferritin H subunit) - Gallus gallus (Chicken), 179 aa.	6..180 4..178	112/175 (64%) 142/175 (81%)	1e-63
Q95MP7	FERRITIN - Canis familiaris (Dog), 183 aa.	6..180 6..180	112/175 (64%) 143/175 (81%)	2e-63

PFam analysis predicts that the NOV90a protein contains the domains shown in the Table 90E.

Table 90E. Domain Analysis of NOV90a			
Pfam Domain	NOV90a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Bacteriofer: domain 1 of 1	14..159	35/172 (20%) 76/172 (44%)	6.7
ferritin: domain 1 of 1	17..173	92/161 (57%) 138/161 (86%)	4.7e-87

Example 91.

The NOV91 clone was analyzed, and the nucleotide and predicted polypeptide

5   sequences are shown in Table 91A.

Table 91A. NOV91 Sequence Analysis			
	SEQ ID NO: 261	487 bp	
NOV91a, CG59710-01 DNA Sequence	TGCTGTGCGTTGTCTTTCCTTCTCACTCAAGCCTGTGAAATCTCTCTTTCAGGTTGAC AGACTAATGGAGTTGCATTTTAAATATCTGGGTGCAATGCAGGTGGCGGACAAGAAGA TTGAAGGGGAAAAACACGACATGGTCCGGCGAGGAGAGATCATCGACAATGACACCGA GGAGGAGTTCTACCTCCGGCGCCTGGATGCGGGGCTCTTTGTTCTCCAGCACATCTGC TACATCATGGCCGAGATCTGCAATGCCAATGTCCCCCAGATTCGCCAGAGGGTTCACC AGATCCTAAACATGCGAGGAAGCTCCATCAAAATTGTCAGGCATATCATCAAGGAGTA TGCAGAGAACATCGGGGACGGCCGGAGCCCGGAGTTCCGGGAGAACGAGCAAAAGCGC ATCCTGGGCTTGCTGGAGAACTTCTAGAGGCACCTTGGCCCTGCGCATCATGGACTCT CTCAGCTTCCCTCCCAGGATCAG		
	ORF Start: ATG at 65	ORF Stop: TAG at 431	
	SEQ ID NO: 262	122 aa	MW at 14385.4kD
NOV91a, CG59710-01 Protein Sequence	MELHFKYLGAMQVADKKIEGEKHDMVRRGEIIDNDTEEEFYLRRLDAGLFVLQHICYI MAEICNANVPQIRQRVHQILNMRGSSIKIVRHIIKEYAENIGDGRSPEFRENEQKRIL GLLENF		

Further analysis of the NOV91a protein yielded the following properties shown in Table 91B.

Table 91B. Protein Sequence Properties NOV91a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV91a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 91C.

Table 91C. Geneseq Results for NOV91a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV91a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU28058	Novel human secretory protein, Seq ID No 227 - Homo sapiens, 518 aa. [WO200166689-A2, 13-SEP-2001]	1..122 397..518	122/122 (100%) 122/122 (100%)	1e-66
AAM93729	Human polypeptide, SEQ ID NO: 3689 - Homo sapiens, 563 aa. [EP1130094-A2, 05-SEP-2001]	1..122 442..563	122/122 (100%) 122/122 (100%)	1e-66
AAB63116	Human secreted protein sequence encoded by gene 39 SEQ ID NO:126 - Homo sapiens, 401 aa. [WO200061748-A1, 19-OCT-2000]	1..119 283..401	119/119 (100%) 119/119 (100%)	1e-64
AAU28246	Novel human secretory protein, Seq ID No 603 - Homo sapiens, 360 aa. [WO200166689-A2, 13-SEP-2001]	1..118 197..316	104/120 (86%) 106/120 (87%)	2e-51
ABB21673	Protein #3672 encoded by probe for measuring heart cell gene expression - Homo sapiens, 32 aa. [WO200157274-A2, 09-AUG-2001]	24..55 1..32	32/32 (100%) 32/32 (100%)	1e-11

In a BLAST search of public sequence databases, the NOV91a protein was found to have homology to the proteins shown in the BLASTP data in Table 91D.

364

Table 91D. Public BLASTP Results for NOV91a				
Protein Accession Number	Protein/Organism/Length	NOV91a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96KD2	TESTES DEVELOPMENT- RELATED NYD-SP19 - Homo sapiens (Human), 376 aa.	1..122 255..376	122/122 (100%) 122/122 (100%)	5e-66
Q9H7A5	CDNA: FLJ21108 FIS, CLONE CAS05257 - Homo sapiens (Human), 225 aa.	1..122 104..225	121/122 (99%) 121/122 (99%)	5e-65
O62703	P14 - Bos taurus (Bovine), 122 aa.	1..122 1..122	116/122 (95%) 119/122 (97%)	2e-62
Q9CWL8	5730471K09RIK PROTEIN - Mus musculus (Mouse), 563 aa.	1..122 442..563	115/122 (94%) 118/122 (96%)	3e-62
Q9Y3M7	DJ633O20.1 (P14L, SIMILAR TO BOS TAURUS P14) - Homo sapiens (Human), 284 aa (fragment).	1..93 192..284	93/93 (100%) 93/93 (100%)	3e-48

PFam analysis predicts that the NOV91a protein contains the domains shown in the Table 91E.

Table 91E. Domain Analysis of NOV91a			
Pfam Domain	NOV91a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 92.

The NOV92 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 92A.

Table 92A. NOV92 Sequence Analysis		
	SEQ ID NO: 263	6527 bp
NOV92a, CG59754-02 DNA Sequence	CCACAGAGGGGAAATGCCAGCTTCCCTCTCCCTGGGGCTCCGTGCCCCTCTGATCCA GCCCTTCGAATTCCCACCCGCCTCCATCGGCCAGCTGCTCTACATTCCCTGTGTGGTG TCCTCGGGGGACATGCCCATCCGTATCACCTGGAGGAAGGACGGACAGGTGATCATCT CAGGCTCGGGCGTGACCATCGAGAGCAAGGAATTCATGAGCTCCCTGCAGATCTCTAG CGTCTCCCTCAAGCACAACGGCAACTATACATGCATCGCCAGCAACGCAGCCGCCACC GTGAGCATTGTGTCTCCAGAACACAGGTTTTTTATTACCTACCACGGCGGGCTGTACA TCTCTGACGTACAGAAGGAGGACGCCCTCTCCACCTATCGCTGCATCACCAAGCACAA GTATAGCGGGGAGACCCGGCAGAGCAATGGGGCACGCCTCTCTGTGACAGACCCTGCT GAGTCGATCCCCACCATCCTGGATGGCTTCCACTCCCAGGAAGTGTGGGCCGGCCACA	

```
CCGTGGAGCTGCCCTGCACCGCCTCGGGCTACCCTATCCCCGCCATCCGCTGGCTCAA
GGATGGCCGGCCCCTCCCGGCTGACAGCCGCTGGACCAAGCGCATCACAGGGCTGACC
ATCAGCGACTTGCGGACCGAGGACAGCGGCACCTACATTTGTGAGGTCACCAACACCT
TCGGTTCGGCAGAGGCCACAGGCATCCTCATGGTCATTGATCCCCTTCATGTGACCCT
GACACCAAAGAAGCTGAAGACCGGCATTGGCAGCACGGTCATCCTCTCCTGTGCCCTG
ACGGGCTCCCCAGAGTTCACCATCCGCTGGTATCGCAACACGGGAGCTGGTGCTGCCTG
ACGAGGCCATCTCCATCCGCGGGCTCAGCAACGAGACGCTGCTCATCACCTCGGCCCA
GAAGAGCCATTCCGGGGCCTACCAGTGCTTCGCTACCCGCAAGGCCCAGACCGCCCAG
GACTTTGCCATCATTGCACTTGAGGATGGCACGCCCCGCATCGTCTCGTCCTTCAGCG
AGAAGGTGGTCAACCCCGGGGAGCAGTTCTCACTGATGTGTGCGGCCAAGGGCGCCCC
GCCCCCCACGGTCACCTGGGCCCTCGACGATGAGCCCATCGTGCGGGATGGCAGCCAC
CGCACCAACCAGTACACCATGTCGGACGGCCACCACCATCAGCCACATGAACGTCACAG
GCCCCCAGATCCGCGACGGGGGCGTGTACCGGTGCACAGCGCGGAACTTGGTGGGCAG
TGCTGAATATCAGGCGCGAATAAACGTAAGAGGCCCACCCAGCATCCGGGCTATGCGG
AACATCACAGCAGTCGCCGGGCGGGACACCCTTATCAACTGCAGGGTCATCGGCTATC
CCTACTACTCCATCAAGTGGTACAAGGATGCCTTGCTGCTGCCAGACAACCACCGCCA
GGTGGTGTTTGAGAATGGGACCCTCAAGCTGACTGACGTGCAGAAGGGCATGGATGAG
GGGGAGTACCTGTGCAGTGTCCTCATCCAGCCCCAGCTCTCCATCAGCCAGAGCGTTC
ACGTAGCCGTCAAAGTGCCCCCTCTGATCCAGCCCTTCGAATTCCCACCCGCCTCCAT
CGGCCAGCTGCTCTACATTCCCTGTGTGGTGTCCTCGGGGGACATGCCCATCCGTATC
ACCTGGAGGAAGGACGGACAGGTGATCATCTCAGGCTCGGGCGTGACCATCGAGAGCA
AGGAATTCATGAGCTCCCTGCAGATCTCTAGCGTCTCCCTCAAGCACAACGGCAACTA
TACATGCATCGCCAGCAACGCAGCCGCCACCGTGAGCCGGGAGCGTCAGCTCATCGTG
CGTGTGCCCCCTCGATTTGTGGTGCAACCCAACAACCAGGATGGCATCTACGGCAAAG
CTGGTGTGCTCAACTGCTCGGTGGACGGCTACCCCCCACCCAAGGTCATGTGGAAGCA
TGCCAAGGGGAGCGGGAACCCCCAGCAGTACCACCCTGTGCCCCTCACTGGCCGCATC
CAGATCCTGCCCAACAGCTCGCTGCTGATCCGCCACGTCCTAGAAGAGGACATCGGCT
ACTACCTCTGCCAGGCCAGCAACGGCGTAGGCACCGACATCAGCAAGTCCATGTTCCT
CACAGTCAAGATCCCGGCCATGATCACTTCCCACCCCAACACCACCATCGCCATCAAG
GGCCATGCGAAGGAGCTAAACTGCACGGCACGGGGTGAGCGGCCCATCATCATCCGCT
GGGAGAAGGGGGACACAGTCATCGACCCTGACCGCGTCATGCGGTATGCCATCGCCAC
CAAGGACAACGGCGACGAGGTCGTCTCCACACTGAAGCTCAAGCCCGCTGACCGTGGG
GACTCTGTGTTCTTCAGCTGCCATGCCATCAACTCGTATGGGGAGGACCGGGGCTTGA
TCCAACTCACTGTGCAAGAGCCCCCCGACCCCCCAGAGCTGGAGATCCGGGAGGTGAA
GGCCCGGAGCATGAACCTGCGCTGGACCCAGCGATTCGACGGGAACAGCATCATCACG
GGCTTCGACATTGAATACAAGAACAAATCAGATTCCTGGGACTTCAAGCAGTCCACAC
GCAACATCTCCCCCACCATCAACCAGGCCAACATTGTGGACTTGCACCCGGCCATCTGT
GTACAGCATCCGCATGTACTCTTTCAACAAGATTGGCCGCAGTGAACCAAGCAAGGAG
CTCACCATCAGCACTGAGGAGGCCGCTCCCGATGGGCCCCCCATGGATGTTACCTTGC
AGCCAGTGACCTCACAGAGCATCCAGGTGACCTGGAAGGCACCCAAGAAGGAGCTGCA
GAACGGTGTCATCCGGGGCTACCAGATTGGCTACAGAGAGAACAGCCCCGGCAGCAAC
GGGCAGTACAGCATCGTGGAGATGAAGGCCACGGGGGACAGCGAGGTCTACACCCTGG
ACAACCTCAAGAAGTTCGCCCAGTATGGGGTGGTGGTCCAAGCCTTCAATCGGGCTGG
CACGGGGCCCTCTTCCAGCGAGATCAATGCCACCACTCTGGAGGATGTGCCCAGCCAG
CCCCCTGAGAACGTCCGGGCCCTGTCCATCACTTCTGACGTGGCCGTCATCTCCTGGT
CAGAGCCCCGCGCAGCACCCTCAATGGCGTCCTCAAAGGCTATCGGGTCATCTTCTG
GTCCCTCTATGTTGATGGGGGAGTGGGGCGAGATGCAGAACATCACCACCACGCGGGAG
CGGGTGGAGCTGCGGGGCATGGAGAAGTTCACCAACTACAGCGTCCAGGTGCTGGCCT
ACACCCAGGCTGGGGACGGCGTACGCAGCAGTGTGCTCTACATCCAGACCAAGGAGGA
CGTTCCAGGTCCCCCTGCTGGCATCAAAGCTGTCCCTTCATCAGCTAGCAGTGTGGTT
GTGTCTTGGCTCCCCCCTACCAAGCCCAACGGGGTGATCCGCAAGTACACCATCTTCT
GTTCCAGCCCCGGGTCTGGCCAGCCGGCTCCCAGCGAGTACGAGACGAGTCCAGAGCA
GCTCTTCTACCGGATCGCCCACCTAAACCGCGGTCAGCAGTATCTGCTGTGGGTGGCC
GCCGTCACCTCTGCCGGCCGGGGCAACAGCAGCGAGGAGGTGACCATCGAGCCTGCTG
GCAAGGCCCCAGCAAAGATCATCTCCTTTGGGGGCACCGTGACAACACCTTGGATGAA
AGATGTTCGGCTGCCTTGCAATTCAGTGGGAGATCCAGCCCCTGCTGTGAAGTGGACC
AAGGACAGTGAAGACTCGGCCATTCCAGTGTCCATGGATGGGCACCGGCTCATCCACA
CCAATGGCACACTGCTGCTGCGTGCAGTGAAGGCTGAGGACTCTGGCTACTACACGTG
CACGGCCACCAACACTGGTGGCTTTGACACCATCATCGTCAACCTTCTGGTGCAAGTT
CCCCCGGACCAGCCCCGCCTCACTGTCTCCAAAACCTCAGCTTCGTCCATCACCCTGA
CCTGGATTCCAGGTGACAATGGGGGCAGCTCCATCCGAGGCTTCGTGCTACAGTACTC
GGTGGACAACAGCGAGGAGTGGAAGGATGTGTTCATCAGCTCCAGCGAGCGCTCCTTC
AAGCTGGACAGCCTCAAGTGTGGCACGTGGTACAAGGTGAAGCTGGCAGCCAAGAACA
GCGTGGGCTCTGGGCGCATCAGCGAGATCATCGAGGCCAAGACCCACGGGCGGGAGCC
CTCCTTCAGCAAAGACCAACACCTCTTCACCCACATCAACTCCACGCATGCTCGGCTT
AACCTGCAGGGCTGGAACAATGGGGGCTGCCCTATCACAGCCATCGTTCTGGAGTACC
GGCCCAAGGGGACCTGGGCCTGGCAGGGCCTCCGGGCCAACAGCTCCGGGGAGGTGTT
TCTGACGGAACTGCGAGAGGCCACGTGGTACGAGCTGCGCATGAGGGCTTGCAACAGT
GCGGGCTGCGGCAATGAAACAGCCCAGTTCGCCACCCTGGACTACGATGGCAGCACCA
TTCCACCCATCAAGTCTGCTCAAGGTGAAGGGGATGATGTGAAGAAGCTGTTCACCAT
CGGCTGCCCTGTCATCCTGGCCACACTGGGGGTGGCACTGCTCTTCATCGTACGCAAG
AAGAGGAAGGAGAAACGGCTGAAGCGACTCCGAGATGCAAAGAGTTTGGCAGAAATGT
TGATAAGCAAGAACAATAGAAGCTTTGACACCCCTGTGAAAGGGCCACCCCAGGGCCC
ACGGCTACACATTGACATCCCCAGGGTCCAGCTGCTCATCGAGGACAAAGAAGGCATC
AAGCAACTGGGAGATGACAAGGCCACCATCCCTGTGACAGATGCTGAGTTCAGCCAAG
CTGTCAACCCACAGAGCTTCTGTACTGGCGTCTCCTTGCACCACCCAACCCTCATCCA
GAGCACAGGACCCCTCATCGACATGTCTGACATCCGGCCAGGAACCAATCCAGTGTCC
```

366

	AGGAAGAATGTGAAGTCAGCCCACAGCACCCGGAACCGGTACTCAAGCCAGTGGACCC TGACCAAGTGCCAGGCCTCCACACCTGCCCGCACCCTCACCTCCGACTGGCGCACCGT GGGCTCCCAGCATGGTGTCACGGTCACTGAGAGTGACAGCTACAGTGCCAGCCTGTCC CAGGACACAGACAAAGGAAGGAACAGCATGGTGTCCACTGAGAGTGCCTCTTCCACCT ACGAGGAGCTGGCCCGGGCCTATGAGCATGCCAAGCTGGAGGAGCAGCTGCAGCACGC CAAGTTTGAGATCACCGAGTGCTTCATCTCTGACAGTTCCTCTGACCAGATGACCACA GGCACCAACGAGAACGCCGACAGCATGACATCCATGAGCACACCCTCAGAGCCTGGCA TCTGCCGCTTTACCGCCTCACCACCCAAGCCCCAGGATGCGGACCGGGGCAAAAACGT GGCTGTGCCCATCCCTCACCGGGCCAACAAGAGTGACTACTGCAACCTGCCCCTGTAT GCCAAGTCAGAGGCCTTCTTTCGAAAGGCAGATGGACGTGAGCCCTGCCCCGTGGTCC CACCCCGTGAGGCCTCCATCCGGAACCTGGCTCGAACCTACCACACCCAGGCTCGCCA CCTGACCCTGGACCCTGCCAGCAAGTCCTTGGGCCTTCCCCACCCAGGGGCCCCCGCT GCCGCCTCCACAGCCACCTTACCTCAGAGGACTCTGGCCATGCCAGCCCCCCCAGCCG GCACAGCCCCCCCAGCCCCCGGCCCCACCCCTGCTGAGCCACCCACCGCCCCCAGCGC TGCCCCTCCGGCCCCCAGCACCGAGCCTCCACGAGCCGGGGGCCCACACACCAAAATG GGGGGCTCCAGGGACTCGCTTCTCGAGATGAGCACATGGGGGTAGGGAGGTCTCAGA AGCAGGGGGCCGGGGCCTACTCCAAATCCTACACCCTGGTGTAGGCCGGCAGGAAGA GCAGCCACGCCTGGGCCGCGCCGCGCCGCAGCCCCACACGCCAGCTCGGCTGTTTTTC TGCATTATTTATATTCAACTGACAGACAAAAACCAACCAACGACAAAACAAAAACCCC CAATCATGAACGCCTGTACATAGAACTCTTTTGTACAAATGAAACTATTTTCTTCTTC TCCATGAAGCCAGGGCACAAAGAATTTGACAGTACAAGTCAAATCCCCCACCCCACAA AATATGTGTGGAGATATATATACATATATAGACAGACAGGAACGCCTCCACGAGCTAT ATATCTATATATTTCTCTCACCCTATTTTGAGACAGAGGCACAAAGACTCAGCAATTT TTTTCCCTCCTCCTCACCTTCCCCCCAGTCTAGGTGGTTTTGACAAAGACCAAAATCC CAACTCAGAGACACTGCATGCGATTTTACTGTTCCAAGAAAACCAGGAGTTGCTTCAA TTTGCAGATGCTTATGTGTTAATACCTTTTTCTATGAAAAAAGACCCAGCGCCGTGTG CAATAAAGGTTATGTTTCCAAAAAAAAGCTT

	ORF Start: ATG at 129	ORF Stop: TAG at 5958	
	SEQ ID NO: 264	1943 aa	MW at 211904.3kD
NOV92a, CG59754-02 Protein Sequence	MPIRITWRKDGQVIISGSGVTIESKEFMSSLQISSVSLKHNGNYTCIASNAAATVSIV SPEHRFFITYHGGLYISDVQKEDALSTYRCITKHKYSGETRQSNGARLSVTDPAESIP TILDGFHSQEVWAGHTVELPCTASGYPIPAIRWLKDGRPLPADSRWTKRITGLTISDL RTEDSGTYICEVTNTFGSAEATGILMVIDPLHVTLTPKKLKTGIGSTVILSCALTGSP EFTIRWYRNTELVLPDEAISIRGLSNETLLITSAQKSHSGAYQCFATRKAQTAQDFAI IALEDGTPRIVSSFSEKVVNPGEQFSLMCAAKGAPPPTVTWALDDEPIVRDGSHRTNQ YTMSDGTTISHMNVTGPQIRDGGVYRCTARNLVGSAEYQARINVRGPPSIRAMRNITA VAGRDTLINCRVIGYPYYSIKWYKDALLLPDNHRQVVFENGTLKLTDVQKGMDEGEYL CSVLIQPQLSISQSVHVAVKVPPLIQPFEFPPASIGQLLYIPCVVSSGDMPIRITWRK DGQVIISGSGVTIESKEFMSSLQISSVSLKHNGNYTCIASNAAATVSRERQLIVRVPP RFVVQPNNQDGIYGKAGVLNCSVDGYPPPKVMWKHAKGSGNPQQYHPVPLTGRIQILP NSSLLIRHVLEEDIGYYLCQASNGVGTDISKSMFLTVKIPAMITSHPNTTIAIKGHAK ELNCTARGERPIIIRWEKGDTVIDPDRVMRYAIATKDNGDEVVSTLKLKPADRGDSVF FSCHAINSYGEDRGLIQLTVQEPPDPPELEIREVKARSMNLRWTQRFDGNSIITGFDI EYKNKSDSWDFKQSTRNISPTINQANIVDLHPASVYSIRMYSFNKIGRSEPSKELTIS TEEAAPDGPPMDVTLQPVTSQSIQVTWKAPKKELQNGVIRGYQIGYRENSPGSNGQYS IVEMKATGDSEVYTLDNLKKFAQYGVVVQAFNRAGTGPSSSEINATTLEDVPSQPPEN VRALSITSDVAVISWSEPPRSTLNGVLKGYRVIFWSLYVDGEWGEMQNITTTRERVEL RGMEKFTNYSVQVLAYTQAGDGVRSSVLYIQTKEDVPGPPAGIKAVPSSASSVVVSWL PPTKPNGVIRKYTIFCSSPGSGQPAPSEYETSPEQLFYRIAHLNRGQQYLLWVAAVTS AGRGNSSEKVTIEPAGKAPAKIISFGGTVTTPWMKDVRLPCNSVGDPAPAVKWTKDSE DSAIPVSMDGHRLIHTNGTLLLRAVKAEDSGYYTCTATNTGGFDTIIVNLLVQVPPDQ PRLTVSKTSASSITLTWIPGDNGGSSIRGFVLQYSVDNSEEWKDVFISSSERSFKLDS LKCGTWYKVKLAAKNSVGSGRISEIIEAKTHGREPSFSKDQHLFTHINSTHARLNLQG WNNGGCPITAIVLEYRPKGTWAWQGLRANSSGEVFLTELREATWYELRMRACNSAGCG NETAQFATLDYDGSTIPPIKSAQGEGDDVKKLFTIGCPVILATLGVALLFIVRKKRKE KRLKRLRDAKSLAEMLISKNNRSFDTPVKGPPQGPRLHIDIPRVQLLIEDKEGIKQLG DDKATIPVTDAEFSQAVNPQSFCTGVSLHHPTLIQSTGPLIDMSDIRPGTNPVSRKNV KSAHSTRNRYSSQWTLTKCQASTPARTLTSDWRTVGSQHGVTVTESDSYSASLSQDTD KGRNSMVSTESASSTYEELARAYEHAKLEEQLQHAKFEITECFISDSSSDQMTTGTNE NADSMTSMSTPSEPGICRFTASPPKPQDADRGKNVAVPIPHRANKSDYCNLPLYAKSE AFFRKADGREPCPVVPPREASIRNLARTYHTQARHLTLDPASKSLGLPHPGAPAAAST ATLPQRTLAMPAPPAGTAPPAPGPTPAEPPTAPSAAPPAPSTEPPRAGGPHTKMGGSR DSLLEMSTSGVGRSQKQGAGAYSKSYTLV

	SEQ ID NO: 265	6049 bp
NOV92b, CG59754-01 DNA Sequence	CCACAGAGGGGAAATGCCAGCTTCCCTCTCCCTGGGGCTCCGTGCCCCCTCTGATCCA GCCCTTCGAATTCCCACCCGCCTCCATCGGCCAGCTGCTCTACATTCCCTGTGTGGTG TCCTCGGGGGACATGCCCATCCGTATCACCTGGAGGAAGGACGGACAGGTGATCATCT CAGGCTCGGGCGTGACCATCGAGAGCAAGGAATTCATGAGCTCCCTGCAGATCTCTAG CGTCTCCCTCAAGCACAACGGCAACTATACATGCATCGCCAGCAACGCAGCCGCCACC GTGAGCATTGTGTCTCCAGAACACAGGTTTTTTATTACCTACCACGGCGGGCTGTACA TCTCTGACGTACAGAAGGAGGACGCCCTCTCCACCTATCGCTGCATCACCAAGCACAA	

367

	GTATAGCGGGGAGACCCGGCAGAGCAATGGGGCACGCCTCTCTGTGACAGACCCTGCT
	GAGTCGATCCCCACCATCCTGGATGGCTTCCACTCCCAGGAAGTGTGGGCCGGCCACA
	CCGTGGAGCTGCCCTGCACCGCCTCGGGCTACCCTATCCCCGCCATCCGCTGGCTCAA
	GGATGGCCGGCCCCTCCCGGCTGACAGCCGCTGGACCAAGCGCATCACAGGGCTGACC
	ATCAGCGACTTGCGGACCGAGGACAGCGGCACCTACATTTGTGAGGTCACCAACACCT
	TCGGTTCGGCAGAGGCCACAGGCATCCTCATGGTCATTGATCCCCTTCATGTGACCCT
	GACACCAAAGAAGCTGAAGACCGGCATTGGCAGCACGGTCATCCTCTCCTGTGCCCTG
	ACGGGCTCCCCAGAGTTCACCATCCGCTGGTATCGCAACACGGAGCTGGTGCTGCCTG
	ACGAGGCCATCTCCATCCGCGGGCTCAGCAACGAGACGCTGCTCATCACCTCGGCCCA
	GAAGAGCCATTCCGGGGCCTACCAGTGCTTCGCTACCCGCAAGGCCCAGACCGCCCAG
	GACTTTGCCATCATTGCACTTGAGGATGGCACGCCCCGCATCGTCTCGTCCTTCAGCG
	AGAAGGTGGTCAACCCCGGGGAGCAGTTCTCACTGATGTGTGCGGCCAAGGGCGCCCC
	GCCCCCCACAGTCACCTGGGCCCTCGACGATGAGCCCATCGTGCGGGATGGCAGCCAC
	CGCACCAACCAGTACACCATGTCGGACGGCACCACCATCAGCCACATGAACGTCACAG
	GCCCCCAGATCCGCGACGGGGGCGTGTACCGGTGCACAGCGCGGAACTTGGTGGGCAG
	TGCTGAATATCAGGCGCGAATAAACGTAAGAGGCCCACCCAGCATCCGGGCTATGCGG
	AACATCACAGCAGTCGCCGGGCGGGACACCCTTATCAACTGCAGGGTCATCGGCTATC
	CCTACTACTCCATCAAGTGGTACAAGGATGCCTTGCTGCTGCCAGACAACCACCGCCA
	GGTGGTGTTTGAGAATGGGACCCTCAAGCTGACTGACGTGCAGAAGGGCATGGATGAG
	GGGGAGTACCTGTGCAGTGTCCTCATCCAGCCCCAGCTCTCCATCAGCCAGAGCGTTC
	ACGTAGCCGTCAAAGTGCCCCCTCTGATCCAGCCCTTCGAATTCCCACCCGCCTCCAT
	CGGCCAGCTGCTCTACATTCCCTGTGTGGTGTCCTCGGGGGACATGCCCATCCGTATC
	ACCTGGAGGAAGGACGGACAGGTGATCATCTCAGGCTCGGGCGTGACCATCGAGAGCA
	AGGAATTCATGAGCTCCCTGCAGATCTCTAGCGTCTCCCTCAAGCACAACGGCAACTA
	TACATGCATCGCCAGCAACGCAGCCGCCACCGTGAGCCGGGAGCGTCAGCTCATCGTG
	CGTGTGCCCCCTCGATTTGTGGTGCAACCCAACAACCAGGATGGCATCTACGGCAAAG
	CTGGTGTGCTCAACTGCTCGGTGGACGGCTACCCCCCACCCAAGGTCATGTGGAAGCA
	TGCCAAGGGTAGCGGGAACCCCCAGCAGTACCACCCTGTGCCCCTCACTGGCCGCATC
	CAGATCCTGCCCAACAGCTCGCTGCTGATCCGCCACGTCCTAGAAGAGGACATCGGCT
	ACTACCTCTGCCAGGCCAGCAACGGCGTAGGCACCGACATCAGCAAGTCCATGTTCCT
	CACAGTCAAGATCCCCACCATCCTGGATGGCTTCCACTCCCAGGAAGTGTGGGCCGGC
	CACACCGTGGAGCTGCCCTGCACCGCCTCGGGCTACCCTATCCCCGCCATCCGCTGGC
	TCAAGGATGGCCGGCCCCTCCCGGCTGACAGCCGCTGGACCAAGCGCGCATCACAGGGCT
	GACCATCAGCGACTTGCGGACCGAGGACAGCGGCACCTACATTTGTGAGGTCACCAAC
	ACCTTCGGTGAGGCCACAGGCATCCTCATGGTCATTGGTGAGGAGCCCCCCGACCCCC
	CAGAGCTGGAGATCCGGGAGGTGAAGGCCCGGAGCATGAACCTGCGCTGGACCCAGCG
	ATTCGACGGGAACAGCATCATCACGGGCTTCGACATTGAATACAAGAACAAAATCAGAT
	TCCTGGGACTTCAAGCAGTCCACACGGCAACATCTCCCCCACCATCAACCAGGCCAACA
	TTGTGGACTTGCACCCGGCATCTGTGTACAGCATCCGCATGTACTCTTTCAACAAGAT
	TGGCCGCAGTGAACCAAGCAAGGAGCTCACCATCAGCACTGAGGAGGCCTCAGCTCCC
	GATGGGCCCCCCATGGATGTTACCTTGCAGCCAGTGACCTCACAGACATCCAGGTGA
	CCTGGAAGCAGGCACCCAAGAAGGAGCTGCAGAACGGTGTCATCCGGGGCTACCAGAT
	TGGCTACAGAGAGAACAGCCCCGGCAGCAACGGGCAGTACAGCATCGTGGAGATGAAG
	GCCACGGGGGACAGCGAGGTCTACACCCTGGACAACCTCAAGAAGTTCGCCCAGTATG
	GGGTGGTGGTCCAGGCCTTCAATCGGGCTGGCACGGGGCCCTCTTCCAGCGAGATCAA
	TGCCACCACTCTGGAGGATGTGCCCAGCCAGCCCCCTGAGAACGTCCGGGCCCTGTCC
	ATCACTTCTGACGTGGCCGTCATCTCCTGGTCAGAGCCCCCGCGCAGCACCCTCAATG
	GCGTCCTCAAAGGCTATCGGGTCATCTTCTGGTCCCTCTATGTTGATGGGGAGTGGGG
	CGAGATGCAGAACATCACCACCACGCGGGAGCGGGTGGAGCTGCGGGGGCATGGAGAAG
	TTCACCAACTACAGCGTCCAGGTGCTGGCCTACACCCAGGCTGGGGACGGCGTACGCA
	GCAGTGTGCTCTACATCCAGACCAAGGAGGACGTTCCAGGTCCCCCTGCTGGCATCAA
	AGCTGTCCCTTCATCAGCTAGCAGTGTGGTTGTGTCTTGGCTCCCCCCTACCAAGCCC
	AACGGGGTGATCCGCAAGTACACCATCTTCTGTTCCAGCCCCGCCCCGCAGGCTCCCA
	GCGAGTACGAGACGAGTCCAGAGCAGCTCTTCTACCGGATCGCCCACCTAAACCGCGG
	TCAGCAGTATCTGCTGTGGGTGGCCGCCGTCACCTCTGCCGGCCGGGGCAACAGCAGC
	GAGAAGGTGACCATCGAGCCTGCTGGCAAGGCCCCAGCAAAGATCATCTCCTTTGGGG
	GCACCGTGACAACACCTTGGATGAAAGATGTTCGGCTGCCTTGCAATTCAGTGGGAGA
	TCCAGCCCCTGCTGTGAAGTGGACCAAGGACAGTGAAGACTCGGCCATTCCAGTGTCC
	ATGGATGGGCACCGGCTCATCCACACCAATGGCACACTGCTGCTGCGTGCAGTGAAGG
	CTGAGGACTCTGGCTACTACACGTGCACGGCCACCAACACTGGTGGCTTTGACACCAT
	CATCGTCAACCTTCTGGTGCAAGTTCCCCCGGACCAGCCCCGCCTCACTGTCTCCAAA
	ACCTCAGCTTCGTCCATCACCCTGACCTGGATTCCAGGTGACAATGGGGGCAGCTCCA
	TCCGAGGTTTTGTGCTACAGTACTCGGTGGACAACAGCGAGGAGTGGAAGGATGTGTT
	CATCAGCTCCAGCGAGCGCTCCTTCAAGCTGGACAGCCTCAAGTGTGGCACGTGGTAC
	AAGGTGAAGCTGGCAGCCAAGAACAGCGTGGGCTCTGGGCGCATCAGCGAGATCATCG
	AGGCCAAGACCCACGGGCGGGAGCCCTCCTTCAGCAAAGACCAACACCTCTTCACCCA
	CATCAACTCCACGCATGCTCGGCTTAACCTGCAGGGCTGGAACAATGGGGGCTGCCCT
	ATCACAGCCATCGTTCTGGAGTACCGGCCCAAGGGGACCTGGGCCTGGCAGGGCCTCC
	GGGCCAACAGCTCCGGGGAGGTGTTTCTGACGGAACTGCGAGAGGCCACGTGGTACGA
	GCTGCGCATGAGGGCTTGCAACAGTGCGGGCTGCGGCAATGAAACAGCCCAGTTCGCC
	ACCCTGGACTACGATGGCAGTACCATTCCACCCATCAAGTCTGCTCAAGGTGAAGGGG
	ATGATGTGAAGAAGCTGTTCACCATCGGCTGCCCTGTCATCCTGGCCACACTGGGGGT
	GGCACTGCTCTTCATCGTACGCAAGAAGAGGAAGGAGAAACGGCTGAAGCGACTCCGA
	GATGCAAAGAGTTTGGCAGAAATGTTGATAAGCAAGAACAATAGAAGCTTTGACACCC
	CTGTGAAAGGGCCACCCCAGGGCCCACGGCTACACATTGACATCCCCAGGGTCCAGCT
	GCTCATCGAGGACAAAGAAGGCATCAAGCAACTGGGTGAGGACAAGGCCACCATCCCT
	GTGACAGATGCTGAGTTCAGCCAAGCTGTCAACCCACAGAGCTTCTGTACTGGCGTCT

368

	CCTTGCACCACCCAACCCTCATCCAGAGCACAGGACCCCTCATCGACATGTCTGACAT CCGGCCAGGAACCGATCCAGTGTCCAGGAAGAATGTGAAGTCAGCCCACAGCACCCGG AACCGGTACTCAAGCCAGTGGACCCTGACCAAGTGCCAGGCCTCCACACCTGCCCGCA CCCTCACCTCCGACTGGCGCACCGTGGGCTCCCAGCATGGTGTCACGGTCACTGAGAG TGACAGCTACAGTGCCAGCCTGTCCCAGGACACAGACAAAGGAAGGAACAGCATGGTG TCCACTGAGAGTGCCTCTTCCACCTACGAGGAGCTGGCCCGGGCCTATGAGCATGCCA AGCTGGAGGAGCAGCTGCAGCACGCCAAGTTTGAGATCACCGAGTGCTTCATCTCTGA CAGTTCCTCTGACCAGATGACCACAGGCACCAACGAGAACGCCGACAGCATGACATCC ATGAGCACACCCTCAGAGCCTGGCATCTGCCGCTTTACCGCCTCACCACCCAAGCCCC AGGATGCGGACCGGCTGCTGATGCTGGTCCCAGGTGCCCACCTCCCTCCTCAGTCCAT CCATGTTGTAGCATATGTCAGAATTTCCTTCTTACTGAACAAGGGTGGGGGAGACCTG GCTTCTGATCTTAGCTCCGGCAGAGCTTGCAGTGAGCCGAGATCACGCGGCACCCGGC CACCAACACTGGTGGCTTTGACACCATCATCGTCAACCTGTGAGGCAGGTGACCCCAG GTGGGGACAGGGATGGAGAAAGGGTAGGGATTCCATCATGCGAGAGGGTCATCGAATG GAAGAAGCCAAACCAAGGGAGAGACAGACCTCTGGAGAAACAGAGGTGCACATGGAAG GGGAGGCAGGAGAGCTGGGGAGTGGGAGTGGGAGTGAGGGTGTGGGAGAGCCAGCACC TTCCCGTCACGGGGGGACTCCCCACACCCCATCACAGGGTCCGCCCTTGTGC**TAA**GGG GTGGTGGCTTTCCCCTCACAGTTCCCCCGGACCAGCCCCGCCTCACTGTCTCCAAAAC CTCAGCTTCGTCCATCACCCTGACCTGGATTCCAGGTGACAATGGGGGCAGCTCCATC CGAGGTGAGGAGGGGTCTGGATGCGGGGGGAAGATAGGGGAAGGAATTCTGGGCCCGGG GCAGGGAAGGGGCTTCA
	ORF Start: ATG at 129 | **ORF Stop: TAA at 5853**
	SEQ ID NO: 266 | 1908 aa | MW at 208575.3kD
NOV92b, CG59754-01 Protein Sequence	MPIRITWRKDGQVIISGSGVTIESKEFMSSLQISSVSLKHNGNYTCIASNAAATVSIV SPEHRFFITYHGGLYISDVQKEDALSTYRCITKHKYSGETRQSNGARLSVTDPAESIP TILDGFHSQEVWAGHTVELPCTASGYPIPAIRWLKDGRPLPADSRWTKRITGLTISDL RTEDSGTYICEVTNTFGSAEATGILMVIDPLHVTLTPKKLKTGIGSTVILSCALTGSP EFTIRWYRNTELVLPDEAISIRGLSNETLLITSAQKSHSGAYQCFATRKAQTAQDFAI IALEDGTPRIVSSFSEKVVNPGEQFSLMCAAKGAPPPTVTWALDDEPIVRDGSHRTNQ YTMSDGTTISHMNVTGPQIRDGGVYRCTARNLVGSAEYQARINVRGPPSIRAMRNITA VAGRDTLINCRVIGYPYYSIKWYKDALLLPDNHRQVVFENGTLKLTDVQKGMDEGEYL CSVLIQPQLSISQSVHVAVKVPPLIQPFEFPPASIGQLLYIPCVVSSGDMPIRITWRK DGQVIISGSGVTIESKEFMSSLQISSVSLKHNGNYTCIASNAAATVSRERQLIVRVPP RFVVQPNNQDGIYGKAGVLNCSVDGYPPPKVMWKHAKGSGNPQQYHPVPLTGRIQILP NSSLLIRHVLEEDIGYYLCQASNGVGTDISKSMFLTVKIPTILDGFHSQEVWAGHTVE LPCTASGYPIPAIRWLKDGRPLPADSRWTKRITGLTISDLRTEDSGTYICEVTNTFGE ATGILMVIGEEPPDPPELEIREVKARSMNLRWTQRFDGNSIITGFDIEYKNKSDSWDF KQSTRNISPTINQANIVDLHPASVYSIRMYSFNKIGRSEPSKELTISTEEASAPDGPP MDVTLQPVTSQSIQVTWKQAPKKELQNGVIRGYQIGYRENSPGSNGQYSIVEMKATGD SEVYTLDNLKKFAQYGVVVQAFNRAGTGPSSSEINATTLEDVPSQPPENVRALSITSD VAVISWSEPPRSTLNGVLKGYRVIFWSLYVDGEWGEMQNITTTRERVELRGMEKFTNY SVQVLAYTQAGDGVRSSVLYIQTKEDVPGPPAGIKAVPSSASSVVVSWLPPTKPNGVI RKYTIFCSSPAPQAPSEYETSPEQLFYRIAHLNRGQQYLLWVAAVTSAGRGNSSEKVT IEPAGKAPAKIISFGGTVTTPWMKDVRLPCNSVGDPAPAVKWTKDSEDSAIPVSMDGH RLIHTNGTLLLRAVKAEDSGYYTCTATNTGGFDTIIVNLLVQVPPDQPRLTVSKTSAS SITLTWIPGDNGGSSIRGFVLQYSVDNSEEWKDVFISSSERSFKLDSLKCGTWYKVKL AAKNSVGSGRISEIIEAKTHGREPSFSKDQHLFTHINSTHARLNLQGWNNGGCPITAI VLEYRPKGTWAWQGLRANSSGEVFLTELREATWYELRMRACNSAGCGNETAQFATLDY DGSTIPPIKSAQGEGDDVKKLFTIGCPVILATLGVALLFIVRKKRKEKRLKRLRDAKS LAEMLISKNNRSFDTPVKGPPQGPRLHIDIPRVQLLIEDKEGIKQLGEDKATIPVTDA EFSQAVNPQSFCTGVSLHHPTLIQSTGPLIDMSDIRPGTDPVSRKNVKSAHSTRNRYS SQWTLTKCQASTPARTLTSDWRTVGSQHGVTVTESDSYSASLSQDTDKGRNSMVSTES ASSTYEELARAYEHAKLEEQLQHAKFEITECFISDSSSDQMTTGTNENADSMTSMSTP SEPGICRFTASPPKPQDADRLLMLVPGAHLPPQSIHVVAYVRISFLLNKGGGDLASDL SSGRACSEPRSRGTRPPTLVALTPSSSTCEAGDPRWGQGWRKGRDSIMREGHRMEEAK PRERQTSGETEVHMEGEAGELGSGSGSEGVGEPAPSRHGGTPHTPSQGPPLC

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 92B.

Table 92B. Comparison of NOV92a against NOV92b.		
Protein Sequence	NOV92a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV92b	1..1771 1..1760	1663/1773 (93%) 1681/1773 (94%)

Further analysis of the NOV92a protein yielded the following properties shown in Table 92C.

Table 92C. Protein Sequence Properties NOV92a	
PSort analysis:	0.7000 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV92a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 92D.

Table 92D. Geneseq Results for NOV92a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV92a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU28091	Novel human secretory protein, Seq ID No 260 - Homo sapiens, 1744 aa. [WO200166689-A2, 13-SEP-2001]	200..1943 1..1744	1744/1744 (100%) 1744/1744 (100%)	0.0
AAM78713	Human protein SEQ ID NO 1375 - Homo sapiens, 1744 aa. [WO200157190-A2, 09-AUG-2001]	200..1943 1..1744	1744/1744 (100%) 1744/1744 (100%)	0.0
AAM39040	Human polypeptide SEQ ID NO 2185 - Homo sapiens, 1744 aa. [WO200153312-A1, 26-JUL-2001]	200..1943 1..1744	1744/1744 (100%) 1744/1744 (100%)	0.0
AAW42086	Human Down syndrome-cell adhesion molecule DS-CAM1 -	44..1778 154..1890	1085/1745 (62%) 1357/1745 (77%)	0.0

370

	[WO9817795-A1, 30-APR-1998]			
AAW42087	Human Down syndrome-cell adhesion molecule DS-CAM2 - Homo sapiens, 1571 aa. [WO9817795-A1, 30-APR-1998]	44..1457 154..1564	890/1416 (62%) 1109/1416 (77%)	0.0

In a BLAST search of public sequence databases, the NOV92a protein was found to have homology to the proteins shown in the BLASTP data in Table 92E.

Table 92E. Public BLASTP Results for NOV92a				
Protein Accession Number	Protein/Organism/Length	NOV92a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAL57166	DOWN SYNDROME CELL ADHESION MOLECULE DSCAML1 - Homo sapiens (Human), 2053 aa.	44..1943 155..2053	1889/1900 (99%) 1892/1900 (99%)	0.0
Q9ULT7	KIAA1132 PROTEIN - Homo sapiens (Human), 1822 aa (fragment).	122..1943 1..1822	1822/1822 (100%) 1822/1822 (100%)	0.0
O60469	Down syndrome cell adhesion molecule precursor (CHD2) - Homo sapiens (Human), 2012 aa.	44..1943 154..2012	1123/1920 (58%) 1410/1920 (72%)	0.0
Q9ERC8	DOWN SYNDROME CELL ADHESION MOLECULE - Mus musculus (Mouse), 2013 aa.	44..1943 154..2013	1119/1921 (58%) 1405/1921 (72%)	0.0
AAL57167	DOWN SYNDROME CELL ADHESION MOLECULE DSCAM - Rattus norvegicus (Rat), 2013 aa.	44..1943 154..2013	1119/1921 (58%) 1405/1921 (72%)	0.0

5   PFam analysis predicts that the NOV92a protein contains the domains shown in the Table 92F.

10

Table 92F. Domain Analysis of NOV92a			
Pfam Domain	**NOV92a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ig: domain 1 of 10	1..48	12/49 (24%) 38/49 (78%)	2.7e-05
ig: domain 2 of 10	72..90	8/19 (42%) 14/19 (74%)	85
ig: domain 3 of 10	130..186	22/60 (37%) 46/60 (77%)	2.1e-14
ig: domain 4 of 10	219..278	16/63 (25%) 44/63 (70%)	4.9e-09
ig: domain 5 of 10	312..377	14/69 (20%) 50/69 (72%)	1.5e-07
ig: domain 6 of 10	409..467	12/61 (20%) 41/61 (67%)	4.8e-05
ig: domain 7 of 10	500..561	17/64 (27%) 49/64 (77%)	3.2e-11
ig: domain 8 of 10	594..659	19/69 (28%) 47/69 (68%)	9.4e-07
ig: domain 9 of 10	693..759	9/70 (13%) 47/70 (67%)	7.9e-06
fn3: domain 1 of 6	777..864	22/89 (25%) 65/89 (73%)	3e-16
fn3: domain 2 of 6	876..968	33/93 (35%) 68/93 (73%)	3.1e-16
fn3: domain 3 of 6	980..1069	26/93 (28%) 69/93 (74%)	2.9e-16
fn3: domain 4 of 6	1081..1167	24/88 (27%) 64/88 (73%)	3.7e-17
ig: domain 10 of 10	1194..1255	17/65 (26%) 46/65 (71%)	4.3e-09
fn3: domain 5 of 6	1274..1357	30/86 (35%) 67/86 (78%)	1.2e-18
fn3: domain 6 of 6	1371..1453	27/86 (31%) 53/86 (62%)	0.045

Example 93.

The NOV93 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 93A.

Table 93A. NOV93 Sequence Analysis		
	SEQ ID NO: 267	1272 bp
NOV93a, CG59800-01 DNA Sequence	AGAGCCTGGTATGCAGGAGGTGCTCTGTAAATACCTGCCCCATACATACCCGCCCCAT ACATACCCACCCCATACATACCCACCCCATACATACCTGCCCTGTCCATACCTGCCCC CTACATACCTGCCCCGTCCATACCTGCCCCCTACATACCTGCCCCGTCCATACCTGCC CCCTACATACCTGCTCTGTCTATACCTGTGGCTAGGACTGTGGCCTTGCTTCTTAGCC GCTCAGAGCCTGCCTCCTCCTCTGCAGAGTGGCGGTGGTAGCAGGGCTTCCCGCGCGC CGATGCTGCTCGTGGCCCTGGTGCTCGGCGCCTACTGCCTCTGCGCCCTCCCCGGCCG CTGCCCGCCGGCCGGCCGCGCCCCCGCCGCGGCCCCCGCGCCCTCCGAGCCGTCCAGC TCCGTCCACCGCCCGGGAGCACCCGGCCTGCCTTTGGCCAGCGGTCCCGGCCGCCGGC GCTTCCCGCAAGCGCTCATCGTTGGCGTGAAGAAGGGCGGCACGCGCGCCCTGCTGGA GTTTCTGCGGCTGCACCCCGACCTCCGCGCGCTGGGCTCTGAGCCCCACTTCTTCGAC AGGTGCCCCGACCGCGGCCTCGCCTGGTCCCGGAGTCTGATGCCCCGAACCCTGGATG GGCAGATCACCATGGAGACGACCCCGGGCTACTTCGTGACGCGAGAGGCCCCCCGCCG CATCCACGCCATGTCCCCGGACACGAAGCTGATCGTGGTGGTGCGGAACCCCGTGACC CGGGCCATCTCCGACTAGGCCCAGACGCTCTCCAAGACCCCGGGCCTGCCCAGCTTCC GCGCCCTGGCCTTCCGCCACGGCCTGGGCCCCGTGGACACAGCCTGGAGCGCCGTCCG CATCGGCCTGTACGCCCAGCACCTGGACCACTGGCTGCGCTACTTCCCCCTGTCCCAC TTCCTGTTCGTCAGCGGGGAGCGTCTGGTCAGCGACCCGGCCGGAGAGGTCGGCCGCG TGCAGGACTTCCTGGGCCTGAAACGGGTCGTCACGGACAAGCACTTCTACTTCAACGC CACCAAGGGCTTCCCCTGCCTCAAGAAGGCCCAGGGCGGCAGCCGTCCCCGCTGCCTG GGCAAGTCCAAGGGCCGGCCCACACCCACGCGTGCCCCAGGCCGTGGTCCGGCGCCTGC AGGAGTTCTACCGGCCCTTCAACCGCAGGTTCTACCAGATGACGGGCCAGGACTTCGG CTGGGGCTGAGCGGCACCCTGGGGATGCTCAGCACCTTGATTGACACCCGCTCG
	ORF Start: GAG at 2	ORF Stop: GGC at 1217
	SEQ ID NO: 268	405 aa	MW at 43994.8kD
NOV93a, CG59800-01 Protein Sequence	MQEVLCKYLPHTYPPHTYPPHTYPPHTYLPCPYLPPTYLPRPYLPPTYLPRPYLPPTY LLCLYLWLGLWPCFLAAQSLPPPLQSGGGSRASRAPMLLVALVLGAYCLCALPGRCPP AARAPAPAPAPSEPSSSVHRPGAPGLPLASGPGRRRFPQALIVGVKKGGTRALLEFLR LHPDXRALGSEXHFFDRCXXXGLXWXRSLMPRTLDGQITMEXTPXYFVTREAPRRIHA MSPDTKLIVVVRNPVTRAISDXXQTLSKTPGLPSFRALAFRHGLGPVDTAWSAVRIGL YAQHLDHWLRYFPLSHFLFVSGERLVSDPAGEVGRVQDFLGLKRVVTDKHFYFNATKG FPCLKKAQGGSRPRCLGKSKGRPHPRVPQAXVRRLQEFYRPFNRRFYQMTGQDFGWG

Further analysis of the NOV93a protein yielded the following properties shown in Table 93B.

Table 93B. Protein Sequence Properties NOV93a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 7 and 8

A search of the NOV93a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 93C.

373

Table 93C. Geneseq Results for NOV93a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV93a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB95507	Human protein sequence SEQ ID NO:18067 - Homo sapiens, 390 aa. [EP1074617-A2, 07-FEB-2001]	31..253 11..237	121/229 (52%) 146/229 (62%)	4e-55
AAY17066	Human 3-OST-3B protein - Homo sapiens, 390 aa. [WO9922005-A2, 06-MAY-1999]	31..253 11..237	121/229 (52%) 146/229 (62%)	4e-55
AAB70115	Human 3-OST-3B - Homo sapiens, 391 aa. [WO200113910-A2, 01-MAR-2001]	31..253 11..238	121/230 (52%) 146/230 (62%)	9e-54
AAB70114	Murine 3-OST-3B - Mus sp, 391 aa. [WO200113910-A2, 01-MAR-2001]	31..253 11..238	119/231 (51%) 147/231 (63%)	2e-51
AAU12275	Human PRO5004 polypeptide sequence - Homo sapiens, 367 aa. [WO200140466-A2, 07-JUN-2001]	86..253 45..214	102/170 (60%) 117/170 (68%)	9e-48

In a BLAST search of public sequence databases, the NOV93a protein was found to have homology to the proteins shown in the BLASTP data in Table 93D.

Table 93D. Public BLASTP Results for NOV93a				
Protein Accession Number	**Protein/Organism/Length**	**NOV93a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96QI5	C439A6.1 (NOVEL PROTEIN SIMILAR TO HEPARAN SULFATE (GLUCOSAMINE) 3-O-SULFOTRANSFERASES) - Homo sapiens (Human), 381 aa (fragment).	85..253 61..229	160/169 (94%) 162/169 (95%)	2e-89
Q96RX7	HEPARAN SULPHATE D-GLUCOSAMINYL 3-O-SULFOTRANSFERASE-3B LIKE - Homo sapiens (Human), 311 aa.	95..253. 1..159	153/159 (96%) 155/159 (97%)	1e-85
Q9Y662	HEPARAN SULFATE D-GLUCOSAMINYL 3-O-	31..253 11..237	121/229 (52%) 146/229 (62%)	1e-54

374

	2.8.2.23) - Homo sapiens (Human), 390 aa.			
Q9QZS6	D-GLYCOSAMINYL 3-O-SULFOTRANSFERASE-3B - Mus musculus (Mouse), 390 aa.	31..253 11..237	119/230 (51%) 147/230 (63%)	3e-52
Q9Y278	HEPARAN SULFATE D-GLUCOSAMINYL 3-O-SULFOTRANSFERASE-2 (EC 2.8.2.23) - Homo sapiens (Human), 367 aa.	86..253 45..214	102/170 (60%) 117/170 (68%)	3e-47

PFam analysis predicts that the NOV93a protein contains the domains shown in the Table 93E.

Table 93E. Domain Analysis of NOV93a			
Pfam Domain	NOV93a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 94.

The NOV94 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 94A.

Table 94A. NOV94 Sequence Analysis		
	SEQ ID NO: 269	2949 bp
NOV94a, CG59761-01 DNA Sequence	GTCCGCCTCCGGGCCGCCGAGCCGCAGCCGCCGAGATGGGGGCCGCCCCGGGCCGCGC CCCCGCCGGGTCCCGCCCGCCGCGCTGCCGCTGAGCGCATGGGCCCGGACCGCGCCGC GCCGCTCCGGGAGCCGGGCCCGGGGTCCCGCCACCACCGCGCGCGGGACAGATTGATT CACTTTGGAGCTGTAAGTACTGATGTATTAGGGTGCAGCGCTCATTGTTCATTGACGC AGAGTCCCAAAATGAATATCCAAGAGCAGGGTTTCCCCTTGGACCTCGGAGCAAGTTT CACCGAAGATGCTCCCCGACCCCCAGTGCCTGGTGAGGAGGGAGAACTGGTGTCCACA GACCCGAGGCCCGCCAGCTACAGTTTCTGCTCCGGGAAGGTGTTGGCATTAAAGGTG AGACTTCGACGGCCACTCCGAGGCGCTCGGATCTGGACCTGGGGTATGAGCCTGAGGG CAGTGCCTCCCCCACCCCACCATACTTGAAGTGGGCTGAGTCACTGCATTCCCTGCTG GATGACCAAGATGGGATAAGCCTGTTCAGGACTTTCCTGAAGCAGGAGGGCTGTGCCG ACTTGCTGGACTTCTGGTTTGCCTGCACTGGCTTCAGGAAGCTGGAGCCCTGTGACTC GAACGAGGAGAAGAGGCTGAAGCTGGCGAGAGCCATCTACCGAAAGTACATTCTTGAT AACAATGGCATCGTGTCCCGGCAGACCAAGCCAGCCACCAAGAGCTTCATAAAGGGCT GCATCATGAAGCAGCTGATCGATCCTGCCATGTTTGACCAGGCCCAGACCGAAATCCA GGCCACTATGGAGGAAAACACCTATCCCTCCTTCCTTAAGTCTGATATTTATTTGGAA TATACGAGGACAGGCTCGGAGAGCCCCAAAGTCTGTAGTGACCAGAGCTCTGGGTCAG GGACAGGGAAGGGCATATCTGGATACCTGCCGACCTTAAATGAAGATGAGGAATGGAA GTGTGACCAGGACATGGATGAGGACGATGGCAGAGACGCTGCTCCCCCCGGAAGACTC CCTCAGAAGCTGCTCCTGGAGACAGCTGCCCCGAGGGTCTCCTCCAGTAGACGGTACA GCGAAGGCAGAGAGTTCAGGTATGGATCCTGGCGGGAGCCAGTCAACCCCTATTATGT CAATGCCGGCTATGCCCTGGCCCCAGCCACCAGTGCCAACGACAGCGAGCAGCAGAGC CTGTCCAGCGATGCAGACACCCTGTCCCTCACGGACAGCAGCAGCGTGGATGGGATCCCCC CATACAGGATCCGTAAGCAGCACCGCAGGGAGATGCAGGAGAGCGTGCAGGTCAATGG GCGGGTGCCCCTACCTCACATTCCCCGCACGTACCGGGTGCCGAAGGAGGTCCGCGTG GAGCCTCAGAAGTTCGCGGAGGAGCTCATCCACCGCCTGGAGGCTGTGCAGCGCACGC GGGAGGCCGAGGAGAAGCTGGAGGAGCGGCGGCTGAAGCGCGTGCGCATGGAGGAGGAAGG TGAGGACGGCGATCCATCATCAGGGCCCCCAGGGCCGTGTCACAAGCTGCCTCCCGCC CCCGCTTGGCACCACTTCCCGCCCCGCCTGTGTTGGACATGGGCTTGTGCCGGGCTCC

	GGGATGCACACGAGGAGAACCCTGAGAGCATCCTGGACGAGCACGTACAGCGTGTGCT GAGGACACCTGGCCGCCAGTCGCCTGGGCCTGGCCATCGCTCCCCGGACAGTGGGCAC GTGGCCAAGATGCCAGTGGCACTGGGGGGTGCCGCCTCGGGGCACGGGAAGCACGTAC CCAAGTCAGGGGCGAAGCTGGACGCGGCCGGCCTGCACCACCACCGACACGTCCACCA CCACGTCCACCACAGCACAGCCCGGCCCAAGGAGCAGGTGGAGGCCGAGGCCACCCGC AGGGCCCAGAGCAGCTTCGCCTGGGGCCTGGAACCACACAGCCATGGGGCAAGGTCCC GAGGCTACTCAGAGAGTGTTGGCGCTGCCCCCAACGCCAGTGATGGCCTCGCCCACAG TGGGAAGGTGGGCGTGGCGTGCAAAAGAAATGCCAAGAAGGCCGAGTCGGGGAAGAGC GCCAGCACCGAGGTGCCAGGTGCCTCGGAGGATGCGGAGAAGAACCAGAAAATCATGC AGTGGATCATTGAGGGGGAAAAGGAGATCAGCAGGCACCGCAGGACCGGCCACGGGTC TTCGGGGACGAGGAAGCCACAGCCCCATGAGAACTCCAGACCCTTGTCCCTTGAGCAC CCCTGGGCCGGCCCTCAGCTCCGGACCTCCGTGCAGCCCTCCCACCTCTTCATCCAAG ACCCCACCATGCCACCCCACCCAGCTCCCAACCCCCTAACCCAGCTGGAGGAGGCGCG CCGACGTCTGGAGGAGGAAGAAAAGAGAGCCAGCCGAGCACCCTCCAAGCAGAGGTAT GTGCAGGAGGTTATGCGGCGGGGACGCGCCTGCGTCAGGCCAGCGTGCGCGCCGGTGC TGCACGTGGTACCAGCCGTGTCGGACATGGAGCTCTCCGGACACAGAGACAAGATCGCA GAGGAAGGTGGGCGGCGGGAGTGCCCAGCCGTGTGACAGCATCGTTGTGGCGTACTAC TTCTGCGGGGAACCCATCCCCTACCGCACCCTGGTGAGGGGCCGCGCTGTCACCCTGG GCCAGTTCAAGGAGCTGCTGACCAAAAAGGGCAGCTACAGATACTACTTCAAGAAAGT GAGCGACGAGTTTGACTGTGGGGTGGTGTTTGAGGAGGTTCGAGAGGACGAGGCCGTC CTGCCCGTCTTTGAGGAGAAGATCATCGGCAAAGTGGAGAAGGTGGACTGA**TAGGCTG** **GTGGGCTGGCCGCTGTGCCAGGCGAGGCCCTTGGCGGGCACGGGTGTCACGGCCAGGC** **AGATGACCTCGTACTCAGGAGCCCGATGGGGAACAGTGTTGGGTGTACC**
	ORF Start: ATG at 97 **ORF Stop: TGA at 2833**
	SEQ ID NO: 270 912 aa MW at 101118.1kD
NOV94a, CG59761-01 Protein Sequence	MGPDRAAPLREPGPGSRHHRARDRLIHFGAVSTDVLGCSAHCSLTQSPKMNIQEQGFP LDLGASFTEDAPRPPVPGEEGELVSTDPRPASYSFCSGKGVGIKGETSTATPRRSDLD LGYEPEGSASPTPPYLKWAESLHSLLDDQDGISLFRTFLKQEGCADLLDFWFACTGFR KLEPCDSNEEKRLKLARAIYRKYILDNNGIVSRQTKPATKSFIKGCIMKQLIDPAMFD QAQTEIQATMEENTYPSFLKSDIYLEYTRTGSESPKVCSDQSSGSGTGKGISGYLPTL NEDEEWKCDQDMDEDDGRDAAPPGRLPQKLLLETAAPRVSSSRRYSEGREFRYGSWRE PVNPYYVNAGYALAPATSANDSEQQSLSSDADTLSLTDSSVDGIPPYRIRKQHRREMQ ESVQVNGRVPLPHIPRTYRVPKEVRVEPQKFAEELIHRLEAVQRTREAEEKLEERLKR VRMEEEGEDGDPSSGPPGPCHKLPPAPAWHHFPPRLCWTWACAGLRDAHEENPESILD EHVQRVLRTPGRQSPGPGHRSPDSGHVAKMPVALGGAASGHGKHVPKSGAKLDAAGLH HHRHVHHHVHHSTARPKEQVEAEATRRAQSSFAWGLEPHSHGARSRGYSESVGAAPNA SDGLAHSGKVGVACKRNAKKAESGKSASTEVPGASEDAEKNQKIMQWIIEGEKEISRH RRTGHGSSGTRKPQPHENSRPLSLEHPWAGPQLRTSVQPSHLFIQDPTMPPHPAPNPL TQLEEARRRLEEEEKRASRAPSKQRYVQEVMRRGRACVRPACAPVLHVVPAVSDMELS ETETRSQRKVGGGSAQPCDSIVVAYYFCGEPIPYRTLVRGRAVTLGQFKELLTKKGSY RYYFKKVSDEFDCGVVFEEVREDEAVLPVFEEKIIGKVEKVD

Further analysis of the NOV94a protein yielded the following properties shown in
Table 94B.

Table 94B. Protein Sequence Properties NOV94a	
PSort analysis:	0.6000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV94a protein against the Geneseq database, a proprietary

database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 94C.

Table 94C. Geneseq Results for NOV94a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV94a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG68175	Wnt signaling protein SEQ ID NO:91 - Homo sapiens, 900 aa. [WO200177327-A1, 18-OCT-2001]	13..912 1..900	898/900 (99%) 898/900 (99%)	0.0
AAW96264	Human axin - Homo sapiens, 900 aa. [WO9902179-A1, 21-JAN-1999]	13..912 1..900	898/900 (99%) 898/900 (99%)	0.0
AAW96265	Murine axin - Mus musculus, 992 aa. [WO9902179-A1, 21-JAN-1999]	6..912 84..992	781/914 (85%) 820/914 (89%)	0.0
AAW93569	Human conductin protein - Homo sapiens, 840 aa. [WO9911780-A2, 11-MAR-1999]	60..912 12..840	378/892 (42%) 506/892 (56%)	e-171
AAW93570	Human conductin protein - Homo sapiens, 840 aa. [WO9911780-A2, 11-MAR-1999]	60..912 12..840	378/892 (42%) 506/892 (56%)	e-171

In a BLAST search of public sequence databases, the NOV94a protein was found to have homology to the proteins shown in the BLASTP data in Table 94D.

Table 94D. Public BLASTP Results for NOV94a				
Protein Accession Number	Protein/Organism/Length	NOV94a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O15169	Axin 1 (Axis inhibition protein 1) (hAxin) - Homo sapiens (Human), 900 aa (fragment).	13..912 1..900	898/900 (99%) 898/900 (99%)	0.0
Q96S28	AXIN - Homo sapiens (Human), 862 aa.	50..912 1..862	858/863 (99%) 858/863 (99%)	0.0
O35625	Axin 1 (Axis inhibition protein 1) (Fused protein) - Mus musculus (Mouse), 992 aa (fragment).	6..912 84..992	781/914 (85%) 820/914 (89%)	0.0
O70239	Axin 1 protein (Axis inhibition protein 1) (rAxin) - Rattus norvegicus (Rat), 893 aa (fragment).	6..912 21..893	756/914 (82%) 793/914 (86%)	0.0

| T08422 | negative regualtor axin [imported] - rat, 832 aa. | 46..912 2..832 | 726/872 (83%) 760/872 (86%) | 0.0 |

PFam analysis predicts that the NOV94a protein contains the domains shown in the Table 94E.

Table 94E. Domain Analysis of NOV94a			
Pfam Domain	NOV94a Match Region	Identities/ Similarities for the Matched Region	Expect Value
RGS: domain 1 of 2	137..198	23/75 (31%) 44/75 (59%)	5.6e-06
RGS: domain 2 of 2	231..260	13/30 (43%) 21/30 (70%)	0.12
TP2: domain 1 of 1	585..709	33/147 (22%) 52/147 (35%)	9.6
DIX: domain 1 of 1	830..912	40/86 (47%) 83/86 (97%)	5.6e-44

Example 95.

The NOV95 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 95A.

Table 95A. NOV95 Sequence Analysis		
	SEQ ID NO: 271	2223 bp
NOV95a, CG59756-01 DNA Sequence	TTGCAGGCATCACCCACGCCCTCTGCACCCACGCTGGAGGACGGGGAGGTTGTCAGGG GCTATGATGAGATGAGTGGGGGCCGCTTCGACTTTGATGATGGAGGGGCGTACTGCGG GGGCTGGGAGGGGGGAAAGGCCCATGGGCATGGACTGTGCACAGGCCCCAAGGGCCAG GGCGAATACTCTGGCTCCTGGAACTTTGGCTTTGAGGTGGCAGGTGTCTACACCTGGC CCAGCGGAAACACCTTTGAGGGATACTGGAGCCAGGGCAAACGGCATGGGCTGGGCAT AGAGACCAAGGGGCGCTGGCTCTACAAGGGCGAGTGGACACATGGCTTCAAGGGACGC TACGGAATCCGGCAGAGCTCAAGCAGCGGTGCCAAGTATGAGGGCACCTGGAACAATG GCCTGCAAGACGGCTATGGCACCGAGACCTATGCTGATGGAGGGACGTACCAAGGCCA GTTCACCAACGGCATGCGCCATGGCTACGGAGTACGCCAGAGCGTGCCCTACGGGATG GCCGTGGTGGTGCGCTCGCCGCTGCGCACGTCGCTGTCGTCCCTGCGCAGCGAGCACA GCAACGGCACGGTGGCCCCGGACTCTCCCGCCTCGCCGGCCTCCGACGGCCCCGCGCT GCCCTCGCCCGCCATCCCGCGTGGCGGCTTCGCGCTCAGCCTCCTGGCCAATGCCGAG GCGGCCGCGCGGGCGCCCAAGGGCGGCGGCCTCTTCCAGCGGGGCGCGCTGCTGGGCA AGCTGCGGCGCGCAGAGTCGCGCACGTCCGTGGGTAGCCAGCGCAGCCGTGTCAGCTT CCTTAAGAGCGACCTCAGCTCGGGCGCCAGCGACGCCGCGTCCACCGCCAGCCTGGGA GAGGCCGCCGAGGGCGCCGACGAGGCCGCACCCTTCGAGGCCGATATCGACGCCACCA CCACCGAGACCTACATGGGCGAGTGGAAGAACGACAAACGCTCGGGCTTCGGCGTGAG CGAACGCTCCAGTGGCCTCCGCTACGAGGGCGAGTGGCTGGACAACCTGCGCCACGGC TATGGCTGCACCACGCTGCCCGACGGCCACCGCGAGGAGGGCAAGTACCGCCACAACG TGCTGGTCAAGGACACCAAGCGCCGCATGCTGCAGCTCAAGAGCAACAAGGTCCGCCA GAAAGTGGAGCACAGTGTGGAGGGTGCCCAGCGCGCCGCTGCTATCGCGCGCCAGAAG GCCGAGATTGCCGCCTCCAGGACAAGCCACGCCAAGGCCAAAGCTGAGGCAGCGGAAC AGGCCGCCCTGGCTGCCAACCAGGAGTCCAACATTGCTCGCACTTTGGCCAGGGAGCT GGCTCCGGACTTCTACCAGCCAGGTCCGGAATATCAGAAGCGCCGGCTGCTGCAGGAG ATCCTGGAGAACTCGGAGAGCCTGCTGGAGCCCCCCGACCGGGGCGCCGGCGCAGCGG GCCTCCCACAGCCGCCCCGCGAGAGCCCGCAGCTGCACGAGCGTGAGACCCCTCGGCC CGAGGGTGGCTCCCCGTCACCGGCCGGGACGCCCCCGCAGCCCAAGCGGCCCAGGCCC GGGGTGTCCAAGGACGGCCTGCTGAGCCCAGGCGCCTGGAACGGCGAGCCCAGCGGTG AGGGCAGCCGGTCAGTCACTCCGTCCGAGGGCGCGGGCCGCCGCAGCCCCGCGCGTCC

378

	AGCCACCGAGCGCATGGCCATCGAGGCTCTGCAGGCACCGCCTGCGCCGTCGCGGGAG CCGGAGGTGGCGCTTTACCAGGGCTACCACAGCTATGCTGTGCGCACCACGCCGCCCG AGCCCCCACCCTTTGAGGACCAGCCCGAGCCCGAGGTCTCCGGGTCCGAGTCCGCGCC CTCGTCCCCGGCCACCGCCCCGCTGCAGGCCCCCACGCTCCGAGGCCCCGAGCCTGCA CGCGAGACCCCCGCCAAGCTGGAGCCCAAGCCCATCATCCCCAAAGCCGAGCCCAGGG CCAAGGCCCGCAAGACTGAGGCTCGAGGGCTGACCAAGGCGGGGGCCAAGAAGAAGGC GCGGAAGGAGGCCGCACTGGCGGCAGAGGCGGAGGTGGAGGTGGAAGAGGTCCCCAAC ACCATCCTCATCTGCATGGTGATCCTGCTGAACATCGGCCTGGCCATCCTCTTTGTTC ACCTCCTGACC**TGA**CCGTCGCTTACCAGGTGCAGCCAGCTGGCTGGAGGAGGGGTTGG GGGGCAGGAGCCCCTGGGG
	ORF Start: ATG at 70	**ORF Stop: TGA at 2158**
	SEQ ID NO: 272	**696 aa**	**MW at 74220.7kD**
NOV95a, CG59756-01 Protein Sequence	MSGGRFDFDDGGAYCGGWEGGKAHGHGLCTGPKGQGEYSGSWNFGFEVAGVYTWPSGN TFEGYWSQGKRHGLGIETKGRWLYKGEWTHGFKGRYGIRQSSSSGAKYEGTWNNGLQD GYGTETYADGGTYQGQFTNGMRHGYGVRQSVPYGMAVVVRSPLRTSLSSLRSEHSNGT VAPDSPASPASDGPALPSPAIPRGGFALSLLANAEAAARAPKGGGLFQRGALLGKLRR AESRTSVGSQRSRVSFLKSDLSSGASDAASTASLGEAAEGADEAAPFEADIDATTTET YMGEWKNDKRSGFGVSERSSGLRYEGEWLDNLRHGYGCTTLPDGHREEGKYRHNVLVK DTKRRMLQLKSNKVRQKVEHSVEGAQRAAAIARQKAEIAASRTSHAKAKAEAAEQAAL AANQESNIARTLARELAPDFYQPGPEYQKRRLLQEILENSESLLEPPDRGAGAAGLPQ PPRESPQLHERETPRPEGGSPSPAGTPPQPKRPRPGVSKDGLLSPGAWNGEPSGEGSR SVTPSEGAGRRSPARPATERMAIEALQAPPAPSREPEVALYQGYHSYAVRTTPPEPPP FEDQPEPEVSGSESAPSSPATAPLQAPTLRGPEPARETPAKLEPKPIIPKAEPRAKAR KTEARGLTKAGAKKKARKEAALAAEAEVEVEEVPNTILICMVILLNIGLAILFVHLLT

Further analysis of the NOV95a protein yielded the following properties shown in Table 95B.

Table 95B. Protein Sequence Properties NOV95a	
PSort analysis:	0.8000 probability located in nucleus; 0.7000 probability located in plasma membrane; 0.3133 probability located in microbody (peroxisome); 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV95a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5 several homologous proteins shown in Table 95C.

Table 95C. Geneseq Results for NOV95a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV95a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM79123	Human protein SEQ ID NO 1785 - Homo sapiens, 628 aa. [WO200157190-A2, 09-AUG-2001]	3..696 4..628	293/704 (41%) 377/704 (52%)	e-127
AAM80107	Human protein SEQ ID NO 3753 - Homo sapiens, 378 aa. [WO200157190-A2, 09-AUG-2001]	283..696 24..378	146/421 (34%) 194/421 (45%)	2e-43

379

ABB21683	Protein #3682 encoded by probe for measuring heart cell gene expression - Homo sapiens, 135 aa. [WO200157274-A2, 09-AUG-2001]	257..389 6..135	78/133 (58%) 104/133 (77%)	7e-42
AAM57089	Human brain expressed single exon probe encoded protein SEQ ID NO: 29194 - Homo sapiens, 135 aa. [WO200157275-A2, 09-AUG-2001]	257..389 6..135	78/133 (58%) 104/133 (77%)	7e-42
AAM17323	Peptide #3757 encoded by probe for measuring cervical gene expression - Homo sapiens, 135 aa. [WO200157278-A2, 09-AUG-2001]	257..389 6..135	78/133 (58%) 104/133 (77%)	7e-42

In a BLAST search of public sequence databases, the NOV95a protein was found to have homology to the proteins shown in the BLASTP data in Table 95D.

Table 95D. Public BLASTP Results for NOV95a				
Protein Accession Number	Protein/Organism/Length	NOV95a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9GKY7	JUNCTOPHILIN TYPE 2 - Oryctolagus cuniculus (Rabbit), 694 aa.	1..696 1..694	644/701 (91%) 662/701 (93%)	0.0
Q9ET79	JUNCTOPHILIN TYPE 2 - Mus musculus (Mouse), 696 aa.	1..696 1..696	608/706 (86%) 644/706 (91%)	0.0
Q9BR39	DJ1108D11.1 (NOVEL PROTEIN SIMILAR TO C. ELEGANS T22C1.7) - Homo sapiens (Human), 552 aa (fragment).	128..672 1..545	544/545 (99%) 544/545 (99%)	0.0
Q9GKY8	MITSUGUMIN72/JUNCTOPHILIN TYPE1 - Oryctolagus cuniculus (Rabbit), 662 aa.	1..696 1..662	364/704 (51%) 468/704 (65%)	0.0
Q9ET80	JUNCTOPHILIN TYPE 1 - Mus musculus (Mouse), 660 aa.	1..696 1..660	371/707 (52%) 469/707 (65%)	0.0

PFam analysis predicts that the NOV95a protein contains the domains shown in the Table 95E.

5

380

Table 95E. Domain Analysis of NOV95a

Pfam Domain	NOV95a Match Region	Identities/ Similarities for the Matched Region	Expect Value
MORN: domain 1 of 7	14..36	10/23 (43%) 13/23 (57%)	1.1
MORN: domain 2 of 7	38..59	9/23 (39%) 15/23 (65%)	0.31
MORN: domain 3 of 7	60..77	8/23 (35%) 15/23 (65%)	3
MORN: domain 4 of 7	106..128	11/23 (48%) 20/23 (87%)	3.7e-06
MORN: domain 5 of 7	129..151	8/23 (35%) 15/23 (65%)	0.027
MORN: domain 6 of 7	291..313	12/23 (52%) 19/23 (83%)	0.00056
MORN: domain 7 of 7	314..336	11/23 (48%) 19/23 (83%)	0.00022

Example 96.

The NOV96 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 96A.

Table 96A. NOV96 Sequence Analysis

	SEQ ID NO: 273	3257 bp
NOV96a, CG59708-01 DNA Sequence	CGTAGGCGCTTCGGCCATGACTGCGGAGCTGCAGCAGGACGACGCGGCCGGCGCGGCA GACGGCCACGGCTCGAGCTGCCAAATGCTGTTAAATCAACTGAGAGAAATCACAGGCA TTCAGGACCCTTCCTTTCTCCATGAAGCTCTGAAGGCCAGTAATGGTGACATTACTCA GGCAGTCAGCCTTCTCACTGATGAGAGAGTTAAGGAGCCCAGTCAAGACACTGTTGCT ACAGAACCATCTGAAGTAGAGGGGAGTGCTGCCAACAAGGAAGTATTAGCAAAAGTTA TAGACCTTACTCATGATAACAAAGATGATCTTCAGGCTGCCATTGCTTTGAGTCTACT GGAGTCTCCCAAAATTCAAGCTGATGGAAGAGATCTTAACAGGATGCATGAAGCAACC TCTGCAGAAACTAAACGCTCAAAGAGAAAACGCTGTGAAGTCTGGGGAGAAAACCCCA ATCCCAATGACTGGAGGAGAGTTGATGGTTGGCCAGTTGGGCTGAAAAATGTTGGCAA TACATGTTGGTTTAGTGCTGTTATTCAGTCTCTCTTTCAATTGCCTGAATTTCGAAGA CTTGTTCTCAGTTATAGTCTGCCACAAAATGTACTTGAAAATTGTCGAAGTCATACAG AAAAGAGAAATATCATGTTTATGCAAGAGCTTCAGTATTTGTTTGCTCTAATGATGGG ATCAAATAGAAAATTTGTAGACCCGTCTGCAGCCCTGGATCTATTAAAGGGAGCATTC CGATCATCTGAGGAACAGCAGCAAGATGTGAGTGAATTCACACACAAGCTCCTGGATT GGCTAGAGGACGCATTCCAGCTAGCTGTTAATGTTAACAGTCCCAGGAACAAATCTGA AAATCCAATGGTGCAGCTGTTCTATGGTACTTTCCTGACTGAAGGGGTTCGTGAAGGA AAACCCTTTTGTAACAATGAGACCTTCGGCCAGTATCCTCTTCAGGTAAACGGTTATC GCAACTTAGACGAGTGTTTGGAAGGGGCCATGGTGGAGGGTGATGTTGAGCTTCTTCC CTCCGATCACTCGGTGAAGTATGGACAAGAGCGTTGGTTTACAAAGCTACCTCCAGTG TTGACCTTTGAACTCTCAAGATTTGAGTTTAATCAGTCCCTTGGGCAGCCAGAGAAAA TTCACAATAAGCTGGAATTTCCTCAGATTATTTATATGGACAGGTACATGTACAGGAG CAAGGAGCTTATTCGAAATAAGAGAGAGTGTATTCGAAAGTTGAAGGAGGAAATAAAA ATTCTGCAGCAAAAATTGGAAAGGTATGTGAAATATGGCTCAGGCCCAGCTCGGTTCC CGCTCCCGGACATGCTGAAATATGTTATTGAATTTGCTAGTACAAAACCTGCCTCAGA AAGCTGTCCACCTGAAAGTGACACACATATGACATTACCACTTTCTTCAGTGCACTGC TCGGTTTCTGACCAGACATCCAAGGAAAGTACAAGTACAGAAAGCTCTTCTCAGGATG

381

	TTGAAAGTACCTTTTCTTCTCCTGAAGATTCTTTACCCAAGTCTAAACCACTGACATC
	TTCTCGGTCTTCCATGGAAATGCCTTCACAGCCAGCTCCACGAACAGTCACAGATGAG
	GAGATAAATTTTGTTAAGACCTGTCTTCAGAGATGGAGGAGTGAGATTGAACAAGATA
	TACAAGATTTAAAGACTTGTATTGCAAGTACTACTCAGACTATTGAACAGATGTACTG
	CGATCCTCTCCTTCGTCAGGTGCCTTATCGCTTGCATGCAGTTCTTGTTCATGAAGGA
	CAAGCAAATGCTGGACACTATTGGGCCTATATCTATAATCAACCCCGACAGAGCTGGC
	TCAAGTACAATGACATCTCTGTTACTGAATCTTCCTGGGAAGAAGTTGAAAGAGATTC
	CTATGGAGGCCTGAGAAATGTTAGTGCTTACTGTCTGATGTACATTAATGCCAAACTA
	CCCTACTTCAATGCAGAGGCAGCCCCAACTGAATCAGATCAAATGTCAGAAGTGGAAG
	CCCTATCTGTGGAACTCAAGCATTACATTCAGGAGGATAACTGGCGGTTTGAGCAGGA
	AGTAGAGGAGTGGGAAGAAGAGCAGTCTTGCAAAATCCCTCAAATGGAGTCCTCCCCC
	AACTCCTCATCACAGGGCTACTCTACATCACAAGAGCCTTCAGTAGCCTCTTCTCATG
	GGGTTCGCTGCTTGTCATCTGAGCATGCTGTGATTGTAAAGGAGCAAACTGCCCAGGC
	TATTGCAAACACAGCCCGTGCCTATGAGAAGAGCGGTGTAGAAGCGGCACTGAGTGAG
	GCATTCCATGAAGAATACTCCAGGCTCTATCAGCTTGCCAAAGAGACCCCCACCTCTC
	ACAGTGATCCTCGACTTCAGCATGTCCTTGTCTACTTTTTCCAAAATGAAGCACCCAA
	AAGGGTAGTAGAACGAACCCTTCTGGAACAGTTTGCAGATAAAAATCTTAGCTATGAT
	GAAAGATCAATCAGCATTATGAAGGTGGCTCAAGCGAAACTGAAGGAAATTGGTCCAG
	ATGACATGAATATGGAAGAGTACAAGAGGTGGCATGAAGATTATAGTTTGTTCCGAAA
	AGTGTCTGTGTATCTCCTAACAGGCCTAGAACTCTATCAAAAAGGAAAGTACCAAGAG
	GCACTTTCCTACCTGGTATATGCCTACCAGAGCAATGCTGCCCTGCTGATGAAGGGGC
	CCCGCCGGGGGGTCAAAGAATCCGTGATTGCTTTATACCGAAGAAAATGCCTTCTGGA
	GCTGAATGCCAAAGCAGCTTCTCTTTTTGAAACAAATGATGATCACTCCGTAACTGAG
	GGCATTAATGTGATGAATGAACTGATCATCCCCTGCATTCACCTTATCATTAATAATG
	ACATTTCCAAGGATGATCTGGATGCCATTGAGGTCATGAGAAACCATTGGTGCTCTTA
	CCTTGGGCAAGATATTGCAGAAAATCTGCAGCTGTGCCTAGGGGAGTTTCTACCCAGA
	CTTCTAGATCCTTCTGCAGAAATCATCGTCTTGAAAGAGCCTCCAACTATTCGACCCA
	ATTCTCCCTATGACCTATGTAGCCGATTTGCAGCTGTCATGGAGTCAATTCAGGGAGT
	TTCAACTGTGACAGTGAAATAAGCTCCCACATGTTCAAGGCCCATTCTGGTTCCTGGC
	TGCCTGCCTCTTGCACAGAAGTTCGTTGTCATAGTGCTCACCTTGGGAAAAGGATTAG
	GTGGGCACA

	ORF Start: ATG at 17	ORF Stop: TAA at 3152
	SEQ ID NO: 274	1045 aa	MW at 119041.7kD

NOV96a, CG59708-01 Protein Sequence	MTAELQQDDAAGAADGHGSSCQMLLNQLREITGIQDPSFLHEALKASNGDITQAVSLL TDERVKEPSQDTVATEPSEVEGSAANKEVLAKVIDLTHDNKDDLQAAIALSLLESPKI QADGRDLNRMHEATSAETKRSKRKRCEVWGENPNPNDWRRVDGWPVGLKNVGNTCWFS AVIQSLFQLPEFRRLVLSYSLPQNVLENCRSHTEKRNIMFMQELQYLFALMMGSNRKF VDPSAALDLLKGAFRSSEEQQQDVSEFTHKLLDWLEDAFQLAVNVNSPRNKSENPMVQ LFYGTFLTEGVREGKPFCNNETFGQYPLQVNGYRNLDECLEGAMVEGDVELLPSDHSV KYGQERWFTKLPPVLTFELSRFEFNQSLGQPEKIHNKLEFPQIIYMDRYMYRSKELIR NKRECIRKLKEEIKILQQKLERYVKYGSGPARFPLPDMLKYVIEFASTKPASESCPPE SDTHMTLPLSSVHCSVSDQTSKESTSTESSSQDVESTFSSPEDSLPKSKPLTSSRSSM EMPSQPAPRTVTDEEINFVKTCLQRWRSEIEQDIQDLKTCIASTTQTIEQMYCDPLLR QVPYRLHAVLVHEGQANAGHYWAYIYNQPRQSWLKYNDISVTESSWEEVERDSYGGLR NVSAYCLMYINAKLPYFNAEAAPTESDQMSEVEALSVELKHYIQEDNWRFEQEVEEWE EEQSCKIPQMESSPNSSSQGYSTSQEPSVASSHGVRCLSSEHAVIVKEQTAQAIANTA RAYEKSGVEAALSEAFHEEYSRLYQLAKETPTSHSDPRLQHVLVYFFQNEAPKRVVER TLLEQFADKNLSYDERSISIMKVAQAKLKEIGPDDMNMEEYKRWHEDYSLFRKVSVYL LTGLELYQKGKYQEALSYLVYAYQSNAALLMKGPRRGVKESVIALYRRKCLLELNAKA ASLFETNDDHSVTEGINVMNELIIPCIHLIINNDISKDDLDAIEVMRNHWCSYLGQDI AENLQLCLGEFLPRLLDPSAEIIVLKEPPTIRPNSPYDLCSRFAAVMESIQGVSTVTV K

	SEQ ID NO: 275	3044 bp

NOV96b, CG59708-02 DNA Sequence	CGTAGGCGCTTCGGCCATGACTGCGGGAGCTGCAGCAGGACGACGCGGCCGGCGCGGCA GACGGCCACGGCTCGAGCTGCCAAATGCTGTTAAATCAACTGAGAGAAATCACAGGCA TTCAGGACCCTTCCTTTCTCCATGAAGCTCTGAAGGCCAGTAATGGTGACATTACTCA GGCAGTCAGCCTTCTCACTGATGAGAGAGTTAAGGAGCCCAGTCAAGACACTGTTGCT ACAGAACCATCTGAAGTAGAGGGGAGTGCTGCCAACAAGGAAGTATTAGCAAAAGTTA TAGACCTTACTCATGATAACAAAGATGATCTTCAGGCTGCCATTGCTTTGAGTCTACT GGAGTCTCCCAAAATTCAAGCTGATGGAAGAGATCTTAACAGGATGCATGAAGCAACC TCTGCAGAAACTAAACGCTCAAAGAGAAATATCATGTTTATGCAAGAGCTTCAGTATT TGTTTGCTCTAATGATGGGATCAAATAGAAAATTTGTAGACCCGTCTGCAGCCCTGGA TCTATTAAAGGGAGCATTCCGATCATCTGAGGAACAGCAGCAAGATGTGAGTGAATTC ACACACAAGCTCCTGGATTGGCTAGAGGACGCATTCCAGCTAGCTGTTAATGTTAACA GTCCCAGGAACAAATCTGAAAATCCAATGGTGCAGCTGTTCTATGGTACTTTCCTGAC TGAAGGGGTTCGTGAAGGAAAACCCTTTTGTAACAATGAGACCTTCGGCCAGTATCCT CTTCAGGTAAACGGTTATCGCAACTTAGACGAGTGTTTGGAAGGGGCCATGGTGGAGG GTGATGTTGAGCTTCTTCCCTCCGATCACTCGGTGAAGTATGGACAAGAGCGTTGGTT TACAAAGCTACCTCCAGTGTTGACCTTTGAACTCTCAAGATTTGAGTTTAATCAGTCC CTTGGGCAGCCAGAGAAAATTCACAATAAGCTGGAATTTCCTCAGATTATTTATATGG ACAGGTACATGTACAGGAGCAAGGAGCTTATTCGAAATAAGAGAGAGTGTATTCGAAA

	GTTGAAGGAGGAAATAAAAATTCTGCAGCAAAAATTGGAAAGGTATGTGAAATATGGC TCAGGCCCAGCTCGGTTCCCGCTCCCGGACATGCTGAAATATGTTATTGAATTTGCTA GTACAAAACCTGCCTCAGAAAGCTGTCCACCTGAAAGTGACACACATATGACATTACC ACTTTCTTCAGTGCACTGCTCGGTTTCTGACCAGACATCCAAGGAAAGTACAAGTACA GAAAGCTCTTCTCAGGATGTTGAAAGTACCTTTTCTTCTCCTGAAGATTCTTTACCCA AGTCTAAACCACTGACATCTTCTCGGTCTTCCATGGAAATGCCTTCACAGCCAGCTCC ACGAACAGTCACAGATGAGGAGATAAATTTTGTTAAGACCTGTCTTCAGAGATGGAGG AGTGAGATTGAACAAGATATACAAGATTTAAAGACTTGTATTGCAAGTACTACTCAGA CTATTGAACAGATGTACTGCGATCCTCTCCTTCGTCAGGTGCCTTATCGCTTGCATGC AGTTCTTGTTCATGAAGGACAAGCAAATGCTGGACACTATTGGGCCTATATCTATAAT CAACCCCGACAGAGCTGGCTCAAGTACAATGACATCTCTGTTACTGAATCTTCCTGGG AAGAAGTTGAAAGAGATTCCTATGGAGGCCTGAGAAATGTTAGTGCTTACTGTCTGAT GTACATTAATGCCAAACTACCCTACTTCAATGCAGAGGCAGCCCCAACTGAATCAGAT CAAATGTCAGAAGTGGAAGCCCTATCGTGGAACTCAAGCATTACATTCAGGAGGATA ACTGGCGGTTTGAGCAGGAAGTAGAGGAGTGGGAAGAAGAGCAGTCTTGCAAAATCCC TCAAATGGAGTCCTCCCCCAACTCCTCATCACAGGGCTACTCTACATCACAAGAGCCT TCAGTAGCCTCTTCTCATGGGGTTCGCTGCTTGTCATCTGAGCATGCTGTGATTGTAA AGGAGCAAACTGCCCAGGCTATTGCAAACACAGCCCGTGCCTATGAGAAGAGCGGTGT AGAAGCGGCACTGAGTGAGGCATTCCATGAGAATACTCCAGGCTCTATCAGCTTGCC AAAGAGACCCCCACCTCTCACAGTGATCCTCGACTTCAGCATGTCCTTGTCTACTTTT TCCAAAATGAAGCACCCAAAAGGGTAGTAGAACGAACCCTTCTGGAACAGTTTGCAGA TAAAAATCTTAGCTATGATGAAAGATCAATCAGCATTATGAAGGTGGCTCAAGCGAAA CTGAAGGAAATTGGTCCAGATGACATGAATATGGAAGAGTACAAGAGGTGGCATGAAG ATTATAGTTTGTTCCGAAAAGTGTCTGTGTATCTCCTAACAGGCCTAGAACTCTATCA AAAAGGAAAGTACCAAGAGGCACTTTCCTACCTGGTATATGCCTACCAGAGCAATGCT GCCCTGCTGATGAAGGGGCCCCGCCGGGGGGGTCAAAGAATCCGTGATTGCTTTATACC GAAGAAAATGCCTTCTGGAGCTGAATGCCAAAGCAGCTTCTCTTTTTGAAACAAATGA TGATCACTCCGTAACTGAGGGCATTAATGTGATGAATGAACTGATCATCCCCTGCATT CACCTTATCATTAATAATGACATTTCCAAGGATGATCTGGATGCCATTGAGGTCATGA GAAACCATTGGTGCTCTTACCTTGGGCAAGATATTGCAGAAATCTGCAGCTGTGCCT AGGGGAGTTTCTACCCAGACTTCTAGATCCTTCTGCAGAAATCATCGTCTTGAAAGAG CCTCCAACTATTCGACCCAATTCTCCCTATGACCTATGTAGCCGATTTGCAGCTGTCA TGGAGTCAATTCAGGGAGTTTCAACTGTGACAGTGAAA**TAA**GCTCCCACATGTTCAAG GCCCATTCTGGTTCCTGGCTGCCTGCCTCTTGCACAGAAGTTCGTTGTCATAGTGCTC ACCTTGGGAAAAGGATTAGGTGGGCACA
	ORF Start: ATG at 17	ORF Stop: TAA at 2939
	SEQ ID NO: 276	974 aa	MW at 110687.3kD
NOV96b, CG59708-02 Protein Sequence	MTAELQQDDAAGAADGHGSSCQMLLNQLREITGIQDPSFLHEALKASNGDITQAVSLL TDERVKEPSQDTVATEPSEVEGSAANKEVLAKVIDLTHDNKDDLQAAIALSLLESPKI QADGRDLNRMHEATSAETKRSKRNIMFMQELQYLFALMMGSNRKFVDPSAALDLLKGA FRSSEEQQQDVSEFTHKLLDWLEDAFQLAVNVNSPRNKSENPMVQLFYGTFLTEGVRE GKPFCNNETFGQYPLQVNGYRNLDECLEGAMVEGDVELLPSDHSVKYGQERWFTKLPP VLTFELSRFEFNQSLGQPEKIHNKLEFPQIIYMDRYMYRSKELIRNKRECIRKLKEEI KILQQKLERYVKYGSGPARFPLPDMLKYVIEFASTKPASESCPPESDTHMTLPLSSVH CSVSDQTSKESTSTESSSQDVESTFSSPEDSLPKSKPLTSSRSSMEMPSQPAPRTVTD EEINFVKTCLQRWRSEIEQDIQDLKTCIASTTQTIEQMYCDPLLRQVPYRLHAVLVHE GQANAGHYWAYIYNQPRQSWLKYNDISVTESSWEEVERDSYGGLRNVSAYCLMYINAK LPYFNAEAAPTESDQMSEVEALSVELKHYIQEDNWRFEQEVEEWEEEQSCKIPQMESS PNSSSQGYSTSQEPSVASSHGVRCLSSEHAVIVKEQTAQAIANTARAYEKSGVEAALS EAFHEEYSRLYQLAKETPTSHSDPRLQHVLVYFFQNEAPKRVVERTLLEQFADKNLSY DERSISIMKVAQAKLKEIGPDDMNMEEYKRWHEDYSLFRKVSVYLLTGLELYQKGKYQ EALSYLVYAYQSNAALLMKGPRRGVKESVIALYRRKCLLELNAKAASLFETNDDHSVT EGINVMNELIIPCIHLIINNDISKDDLDAIEVMRNHWCSYLGQDIAENLQLCLGEFLP RLLDPSAEIIVLKEPPTIRPNSPYDLCSRFAAVMESIQGVSTVTVK
	SEQ ID NO: 277	3231 bp
NOV96c, CG59708-03 DNA Sequence	GCGCTTCGGCC**ATG**ACTGCGGAGCTGCAGCAGGACGACGCGGCCGGCGCGGCAGACGG CCACGGCTCGAGCTGCCAAATGCTGTTAAATCAACTGAGAGAAATCACAGGCATTCAG GACCCTTCCTTTCTCCATGAAGCTCTGAGGGCCAGTAATGGTGACATTACTCAGGCAG TCAGCCTTCTCACTGATGAGAGAGTTAAGGAGCCCAGTCAAGACACTGTTGCTACAGA ACCATCTGAAGTAGAGGGGAGTGCTGCCAACAAGGAAGTATTAGCAAAAGTTATAGAC CTTACTCATGATAACAAAGATGATCTTCAGGCTGCCATTGCTTTGAGTCTACTGGAGT CTCCCAAAATTCAAGCTGATGGAAGAGATCTTAACAGGATGCATGAAGCAACCTCTGC AGAAACTAAACGCTCAAAGAGAAAACGCTGTGAAGTCTGGGGAGAAAACCCAATCCC AATGACTGGAGGAGAGTTGATGGTTGGCCAGTTGGGCTGAAAAATGTTGGCAATACAT GTTGGTTTAGTGCTGTTATTCAGTCTCTCTTTCAATTGCCTGAATTTCGAAGACTTGT TCTCAGTTATAGTCTGCCACAAAATGTACTTGAAAATTGTCGAAGTCATACAGAAAAG AGAAATATCATGTTTATGCAAGAGCTTCAGTATTTGTTTGCTCTAATGATGGGATCAA ATAGAAAATTTGTAGACCCGTCTGCAGCCCTGGATCTATTAAAGGGAGCATTCCGATC ATCTGAGGAACAGCAGCAAGATGTGAGTGAATTCACACACAAGCTCCTGGATTGGCTA GAGGACGCATTCCAGCTAGCTGTTAATGTTAACAGTCCCAGGAACAAATTTGAAAATC CAATGGTGCAGCTGTTCTATGGTACTTTCCTGACTGAAGGGGGTTCGTGAAGGAAAACC

	CTTTTGTAACAATGAGACCTTCGGCCAGTATCCTCTTCAGGTAAACGGTTATCGCAAC
	TTAGACGAGTGTTTGGAAGGGGCCATGGTGGAGGGTGATGTTGAGCTTCTTCCCTCCG
	ATCACTCGGTGAAGTATGGACAAGAGCGTTGGTTTACAAAGCTACCTCCAGTGTTGAC
	CTTTGAACTCTCAAGATTTGAGTTTAATCAGTCCCTTGGGCAGCCAGAGAAAATTCAC
	AATAAGCTGGAATTTCCTCAGATTATTTATATGGACAGGTACATGTACAGGAGCAAGG
	AGCTTATTCGAAATAAGAGAGAGTGTATTCGAAAGTTGAAGGAGGAAATAAAAATTCT
	GCAGCAAAAATTGGAAGGGTATGTGAAATATGGCTCAGGCCCAGCTCGGTTCCCGCTC
	CCGGACATGCTGAAATATGTTATTGAATTTGCTAGTACAAAACCTGCCTCAGAAAGCT
	GTCCACCTGAAAGTGACACACATATGACATTACCACTTTCTTCAGTGCACTGCTCGGT
	TTCTAACCAGACATCCAAGGAAAGTACAAGTACAGAAAGCTCTTCTCAGGATGTTGAA
	AGTACCTTTTCTTCTCCTGAAGATTCTTTACCCAAGTCTAAACCACTGACATCTTCTC
	GGTCTTCCATGGAAATGCCTTCACAGCCAGCTCCACGAACAGTCACAGATGAGGAGAT
	AAATTTTGTTAAGACCTGTCTTCAGAGATGGAGGAGTGAGATTGAACAAGATATACAA
	GATTTAAAGACTTGTATTGCAAGTACTACTCAGACTATTGAACAGATGTACTGCGATC
	CTCTCCTTCGTCAGGTGCCTTATCGCTTGCATGCAGTTCTTGTTCATGAAGGACAAGC
	AAATGCTGGACACTATTGGGCCTATATCTATAATCAACCCCGACAGAGCTGGCTCAAG
	TACAATGACATCTCTGTTACTGAATCTTCCTGGGAAGAAGTTGAAAGAGATTCCTATG
	GAGGCCTGAGAAATGTTAGTGCTTACTGTCTGATGTACATTAACGACAAACTACCCTA
	CTTCAATGCAGAGGCAGCCCCAACTGAATCAGATCAAATGTCAGAAGTGGAAGCCCTA
	TCTGTGGAACTCAAGCATTACATTCAGGAGGATAACTGGCGGTTTGAGCAGGAAGTAG
	AGGAGTGGGAAGAAGAGCAGTCTTGCAAAATCCCTCAAATGGAGTCCTCCACCAACTC
	CTCATCACAGGACTACTCTACATCACAAGAGCCTTCAGTAGCCTCTTCTCATGGGGTT
	CGCTGCTTGTCATCTGAGCATGCTGTGATTGTAAAGGAGCAAACTGCCCAGGCTATTG
	CAAACACAGCCCGTGCCTATGAGAAGAGCGGTGTAGAAGCGGCACTGAGTGAGGCATT
	CCATGAAGAATACTCCAGGCTCTATCAGCTTGCCAAAGAGACCCCCACCTCTCACAGT
	GATCCTCGACTTCAGCATGTCCTTGTCTACTTTTTCCAAAATGAAGCACCCAAAAGGG
	TAGTAGAACGAACCCTTCTGGAACAGTTTGCAGATAAAAATCTTAGCTATGATGAAAG
	ATCAATCAGCATTATGAAGGTGGCTCAAGCGAAACTGAAGGAAATTGGTCCAGATGAC
	ATGAATATGGAAGAGTACAAGAAGTGGCATGAAGATTATAGTTTGTTCCGAAAAGTGT
	CTGTGTATCTCCTAACAGGCCTAGAACTCTATCAAAAAGGAAAGTACCAAGAGGCACT
	TTCCTACCTGGTATATGCCTACCAGAGCAATGCTGCCCTGCTGATGAAGGGGCCCCGC
	CGGGGGGTCAAAGAATCCGTGATTGCTTTATACCGAAGAAAATGCCTTCTGGAGCTGA
	ATGCCAAAGCAGCTTCTCTTTTTGAAACAAATGATGATCACTCCGTAACTGAGGGCAT
	TAATGTGATGAATGAACTGATCATCCCCTGCATTCACCTTATCATTAATAATGACATT
	TCCAAGGATGATCTCGGATGCCATTGAGGTCATGAGAAACCATTGGTGCTCTTACCTTG
	GGCAAGATATTGCAGAAAATCTGCAGCTGTGCCTAGGGGAGTTTCTACCCAGACTTCT
	AGATCCTTCTGCAGAAATCATCGTCTTGAAAGAGCCTCCAACTATTCGACCCAATTCT
	CCCTATGACCTATGTAGCCGATTTGCAGCTGTCATGGAGTCAATTCAGGGAGTTTCAA
	CTGTGACAGTGAAA**TAA**GCTCCCACATGTTCAAGGCCCATTCTGGTTCCTGGCTGCCT
	GCCTCTTGCACAGAAGTTCGTTGTCATAGTGCTCACCTTGG

	ORF Start: ATG at 12	ORF Stop: TAA at 3147	
	SEQ ID NO: 278	1045 aa	MW at 119107.7kD
NOV96c, CG59708-03 Protein Sequence	MTAELQQDDAAGAADGHGSSCQMLLNQLREITGIQDPSFLHEALRASNGDITQAVSLL TDERVKEPSQDTVATEPSEVEGSAANKEVLAKVIDLTHDNKDDLQAAIALSLLESPKI QADGRDLNRMHEATSAETKRSKRKRCEVWGENPNPNDWRRVDGWPVGLKNVGNTCWFS AVIQSLFQLPEFRRLVLSYSLPQNVLENCRSHTEKRNIMFMQELQYLFALMMGSNRKF VDPSAALDLLKGAFRSSEEQQQDVSEFTHKLLDWLEDAFQLAVNVNSPRNKFENPMVQ LFYGTFLTEGVREGKPFCNNETFGQYPLQVNGYRNLDECLEGAMVEGDVELLPSDHSV KYGQERWFTKLPPVLTFELSRFEFNQSLGQPEKIHNKLEFPQIIYMDRYMYRSKELIR NKRECIRKLKEEIKILQQKLEGYVKYGSGPARFPLPDMLKYVIEFASTKPASESCPPE SDTHMTLPLSSVHCSVSNQTSKESTSTESSSQDVESTFSSPEDSLPKSKPLTSSRSSM EMPSQPAPRTVTDEEINFVKTCLQRWRSEIEQDIQDLKTCIASTTQTIEQMYCDPLLR QVPYRLHAVLVHEGQANAGHYWAYIYNQPRQSWLKYNDISVTESSWEEVERDSYGGLR NVSAYCLMYINDKLPYFNAEAAPTESDQMSEVEALSVELKHYIQEDNWRFEQEVEEWE EEQSCKIPQMESSTNSSSQDYSTSQEPSVASSHGVRCLSSEHAVIVKEQTAQAIANTA RAYEKSGVEAALSEAFHEEYSRLYQLAKETPTSHSDPRLQHVLVYFFQNEAPKRVVER TLLEQFADKNLSYDERSISIMKVAQAKLKEIGPDDMNMEEYKKWHEDYSLFRKVSVYL LTGLELYQKGKYQEALSYLVYAYQSNAALLMKGPRRGVKESVIALYRRKCLLELNAKA ASLFETNDDHSVTEGINVMNELIIPCIHLIINNDISKDDLDAIEVMRNHWCSYLGQDI AENLQLCLGEFLPRLLDPSAEIIVLKEPPTIRPNSPYDLCSRFAAVMESIQGVSTVTV K

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 96B.

Table 96B. Comparison of NOV96a against NOV96b through NOV96c.		
Protein Sequence	**NOV96a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV96b	209..1045 138..974	805/837 (96%) 805/837 (96%)
NOV96c	1..1045 1..1045	979/1045 (93%) 981/1045 (93%)

Further analysis of the NOV96a protein yielded the following properties shown in Table 96C.

Table 96C. Protein Sequence Properties NOV96a	
PSort analysis:	0.8800 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV96a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5 several homologous proteins shown in Table 96D.

Table 96D. Geneseq Results for NOV96a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV96a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE04874	Human protease protein-1 (PRTS-1) - Homo sapiens, 1055 aa. [WO200146443-A2, 28-JUN-2001]	22..1036 18..1047	524/1035 (50%) 713/1035 (68%)	0.0
AAB31552	A human ubiquitin specific protease 25 (USP25) - Homo sapiens, 1055 aa. [WO200079267-A2, 28-DEC-2000]	22..1036 18..1047	524/1035 (50%) 713/1035 (68%)	0.0
AAB31546	A human ubiquitin specific protease 25 (USP25) - Homo sapiens, 1055 aa. [WO200078934-A2, 28-DEC-2000]	22..1036 18..1047	524/1035 (50%) 713/1035 (68%)	0.0
AAB74491	Human SYK kinase binding protein	22..1036 18..1047	522/1035 (50%) 710/1035 (68%)	0.0

	sapiens, 1055 aa. [WO200121654-A2, 29-MAR-2001]			
AAB31556	A human ubiquitin specific protease (USP) - Homo sapiens, 1087 aa. [WO200079267-A2, 28-DEC-2000]	22..1036 18..1079	525/1067 (49%) 717/1067 (66%)	0.0

In a BLAST search of public sequence databases, the NOV96a protein was found to have homology to the proteins shown in the BLASTP data in Table 96E.

Table 96E. Public BLASTP Results for NOV96a				
Protein Accession Number	Protein/Organism/Length	NOV96a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96RU2	UBIQUITIN SPECIFIC PROTEASE - Homo sapiens (Human), 1077 aa.	1..1045 1..1077	1041/1077 (96%) 1042/1077 (96%)	0.0
Q9P213	KIAA1515 PROTEIN - Homo sapiens (Human), 757 aa (fragment).	304..1045 16..757	738/742 (99%) 739/742 (99%)	0.0
P57080	Ubiquitin carboxyl-terminal hydrolase 25 (EC 3.1.2.15) (Ubiquitin thiolesterase 25) (Ubiquitin-specific processing protease 25) (Deubiquitinating enzyme 25) (mUSP25) - Mus musculus (Mouse), 1055 aa.	22..1036 18..1047	527/1033 (51%) 710/1033 (68%)	0.0
Q9UHP3	Ubiquitin carboxyl-terminal hydrolase 25 (EC 3.1.2.15) (Ubiquitin thiolesterase 25) (Ubiquitin-specific processing protease 25) (Deubiquitinating enzyme 25) (USP on chromosome 21) - Homo sapiens (Human), 1087 aa.	22..1036 18..1079	525/1067 (49%) 717/1067 (66%)	0.0
Q9H9W1	CDNA FLJ12512 FIS, CLONE NT2RM2001730, WEAKLY SIMILAR TO PROBABLE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE K02C4.3 (EC 3.1.2.15) - Homo sapiens (Human), 737 aa.	313..1036 2..729	363/733 (49%) 510/733 (69%)	0.0

PFam analysis predicts that the NOV96a protein contains the domains shown in the Table 96F.

<table>
<tr><th colspan="4">Table 96F. Domain Analysis of NOV96a</th></tr>
<tr><th>Pfam Domain</th><th>NOV96a Match Region</th><th>Identities/ Similarities for the Matched Region</th><th>Expect Value</th></tr>
<tr><td>UIM: domain 1 of 1</td><td>96..113</td><td>9/18 (50%)
14/18 (78%)</td><td>8.4</td></tr>
<tr><td>UCH-1: domain 1 of 1</td><td>162..193</td><td>14/32 (44%)
28/32 (88%)</td><td>2.6e-11</td></tr>
<tr><td>UCH-2: domain 1 of 1</td><td>580..649</td><td>26/72 (36%)
56/72 (78%)</td><td>1.5e-19</td></tr>
</table>

Example 97.

5   The NOV97 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 97A.

<table>
<tr><th colspan="3">Table 97A. NOV97 Sequence Analysis</th></tr>
<tr><td></td><td>SEQ ID NO: 279</td><td>1601 bp</td></tr>
<tr><td>NOV97a,
CG59559-01 DNA Sequence</td><td colspan="2">AGGGCAGAGGCCACAGCGCCATCCCCTTCCCCATGGTCTCCCTACCCCCAACCTGCAC
TGGGCGCTCCGCCCAGAGGTGAGTCCCTCCCAGCCCTTCTCTCCTTCTGTCCTAGCCA
TCCGCAGAGCCATCCTGTGCAAAGGAAGGAGCTAGGCTGTGCGCCCTGGGCGTCATGA
TCCTTCTGCGGGCCTCCGAAGTGCGGCAGCTGCTTCACAATAAGTTCGTGGTCATCCT
GGGGGACTCTGTGCATAGGGCAGTATACAAGGACCTGGTGCTTCTGCTGCAGAAGGAC
CGCCTGCTCACTCCCGGGCAGCTTAGAGCAAGGGGGGAGCTGAACTTCGAACAAGATG
AGCTGGTGGACGGAGGCCAGCGGGGCCACATGCACAACGGCCTTAACTACCGTGAGGT
CCGCGAGTTCCGCTCCGACCACCATCTGGTACGTTTTTACTTCCTCACCCGCGTGTAC
TCCGATTACCTCCAGACCATCTTGAAAGAGCTGCAGTCGGGCGAGCACGCCCCCGACC
TGGTCATCATGAATTCCTGCCTCTGGGACATCTCCAGGTATGGTCCGAACTCCTGGAG
AAGCTACCTGGAGAACCTGGAGAACCTGTTCCAGTGCCTGGGCCAGGTGCTGCCCGAG
TCTTGCCTCCTGGTGTGGAACACGGCCATGCCTGTGGGCGAGGAAGTCACCGGGGGTT
TTCTTCCGCCCAAGCTCCGGCGGCAGAAGGCCACCTTCCTGAAAAACGAAGTGGTCAA
AGCCAACTTCCACAGCGCCACCGAGGCACGTAAACATAACTTCGATGTACTGGACTTG
CATTTCCACTTCCGCCACGCGAGGGAGAACCTGCACTGGGACGGGGTGCACTGGAATG
GACGTGTGCACCGCTGCCTCTCCCAGCTGCTGCTGGCCCACGTGGCCGACGCCTGGGG
TGTGGAGCTGCCCCACCGCCACCCCGTGGGCGAGTGGATCAAGAAGAAAAAACCTGGC
CCGAGAGTCGAAGGGCCGCCCCAGGCCAACAGAAATCACCCGGCCTTACCTCTGTCCC
CACCCTTACCTTCCCCCACATACCGCCCCCTGCTTGGGTTCCCACCCCAGCGCTTGCC
GCTGCTCCCGCTCCTGTCCCCACAGCCTCCTCCTCCCATTCTCCATCACCAGGGAATG
CCCCGGTTCCCACAGGGTCCCCCAGATGCCTGTTTTTCCTCAGACCATACTTTCCAGT
CGGATCAATTCTATTGCCATTCAGATGTCCCCTCATCAGCCCATGCAGGTTTCTTCGT
CGAAGACAATTTTATGGTTGGTCCTCAGCTGCCTATGCCCTTCTTCCCCACACCCCGT
TATCAGCGGCCTGCCCCAGTGGTACATAGGGGTTTTGGCAGGTATCGTCCCCGTGGCC
CCTATACGCCCTGGGGACAGCGGCCTCGACCTTCAAAGAGAAGGGCCCCAGCCAATCC
TGAGCCAAGGCCTCAATAGACGGACCTAGGCCTTATTTCCTCTTTATGAACATGGATT
GGACAGATCTGACACTTCCTTTCCATTGCTTGGCCTGAACAGACTGACCTTGTTAACT
TAAGCCTGGAGTCCATGCCTCGTCTTCCTTTTGTT</td></tr>
<tr><td></td><td>ORF Start: ATG at 171</td><td>ORF Stop: TAG at 1467</td></tr>
<tr><td></td><td>SEQ ID NO: 280</td><td>432 aa | MW at 49726.6kD</td></tr>
<tr><td>NOV97a,</td><td colspan="2">MILLRASEVRQLLHNKFVVILGDSVHRAVYKDLVLLLQKDRLLTPGQLRARGELNFEQ
DELVDGGQRGHMHNGLNYREVREFRSDHHLVRFYFLTRVYSDYLQTILKELQSGEHAP</td></tr>
</table>

| CG59559-01 Protein Sequence | DLVIMNSCLWDISRYGPNSWRSYLENLENLFQCLGQVLPESCLLVWNTAMPVGEEVTG GFLPPKLRRQKATFLKNEVVKANFHSATEARKHNFDVLDLHFHFRHARENLHWDGVHW NGRVHRCLSQLLLAHVADAWGVELPHRHPVGEWIKKKKPGPRVEGPPQANRNHPALPL SPPLPSPTYRPLLGFPPQRLPLLPLLSPQPPPPILHHQGMPRFPQGPPDACFSSDHTF QSDQFYCHSDVPSSAHAGFFVEDNFMVGPQLPMPFFPTPRYQRPAPVVHRGFGRYRPR GPYTPWGQRPRPSKRRAPANPEPRPQ |

Further analysis of the NOV97a protein yielded the following properties shown in Table 97B.

Table 97B. Protein Sequence Properties NOV97a	
PSort analysis:	0.5937 probability located in mitochondrial matrix space; 0.5103 probability located in microbody (peroxisome); 0.4900 probability located in nucleus; 0.3252 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV97a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 97C.

Table 97C. Geneseq Results for NOV97a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV97a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG74241	Human colon cancer antigen protein SEQ ID NO:5005 - Homo sapiens, 281 aa. [WO200122920-A2, 05-APR-2001]	34..294 1..266	162/268 (60%) 191/268 (70%)	1e-82
AAE03639	Human extracellular matrix and cell adhesion molecule-3 (XMAD-3) - Homo sapiens, 386 aa. [WO200142285-A2, 14-JUN-2001]	1..421 1..366	197/435 (45%) 231/435 (52%)	2e-82

In a BLAST search of public sequence databases, the NOV97a protein was found to have homology to the proteins shown in the BLASTP data in Table 97D.

Table 97D. Public BLASTP Results for NOV97a					
Protein Accession Number	Protein/Organism/Length	NOV97a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96HM7	SIMILAR TO HYPOTHETICAL PROTEIN FLJ22376 - Homo sapiens (Human), 432 aa.	1..432 1..432	432/432 (100%) 432/432 (100%)	0.0
Q96B20	HYPOTHETICAL 31.4 KDA PROTEIN - Homo sapiens (Human), 279 aa.	121..310 1..190	190/190 (100%) 190/190 (100%)	e-116
Q9H1Q7	BA12M19.1.3 (NOVEL PROTEIN) (CDNA FLJ31791 FIS, CLONE NT2RI2008749, WEAKLY SIMILAR TO SPLICEOSOME ASSOCIATED PROTEIN 49) - Homo sapiens (Human), 454 aa.	1..421 18..434	234/437 (53%) 273/437 (61%)	e-111
Q9H1Q6	BA12M19.1.1 (NOVEL PROTEIN) - Homo sapiens (Human), 403 aa.	1..421 18..383	197/435 (45%) 231/435 (52%)	7e-82
Q9H6D1	CDNA: FLJ22376 FIS, CLONE HRC07327 - Homo sapiens (Human), 403 aa.	1..421 18..383	196/435 (45%) 231/435 (53%)	1e-81

PFam analysis predicts that the NOV97a protein contains the domains shown in the Table 97E.

Table 97E. Domain Analysis of NOV97a			
Pfam Domain	NOV97a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 98.

The NOV98 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 98A.

389

<table>
<tr><td colspan="3">Table 98A. NOV98 Sequence Analysis</td></tr>
</table>

	SEQ ID NO: 281	981 bp
NOV98a, CG59669-01 DNA Sequence	GCGCCGGGTCCCAGAATCTAGTCCTACGCCACGGTTTTGACCACGCGTGACCCGCTGC CCAGCCGGCCCGGCCATCAGGTGGTCCGTGTGTCCCTCTGAC**ATG**TCGTCCTGCAGCC GCGTGGCCCTGGTAACTGGGGCTAACAAAGGCATCGGCTTTGCGATCACGCGTGACCT GTGTCGGAAATTCTCCGGGGACGTGGTGCTCACGGCGCGGGACGAGGCGCGGGGCCGC GCGGCGGTGCAGCAGCTGCAGGCGGAGGGCCTGAGCCCACGCTTCCACCAGCTGGACA TCGACGACCCGCAGAGCATCCGTGCGCTGCGCGACTTTCTGCGCAAGGAGTACGGGGG ACTTAACGTGCTGGTCAACAACGCGGGCATCGCCTTTAGAAGTACTGATCTCACCCAC TTTCACATTCTAAGAGAAGCTGCAATGAAAACTAACTTTTTTGGTACCCAGGCCGTCT GCACAGAGCTACTCCCTCTAATAAAAACCCAAGGTAGAGTGGTGAATATATCAAGCCT AATAAGTCTAGAGGCCCTGAAAAACTGCAGCCTGGAGCTACAGCAGAAGTTTCGAAGT GAGACCATCACAGAGGAGGAGCTGGTGGGGCTCATGAACAAGTTTGTGGAGGATACAA AGAAAGGAGTCCATGCAAAAGAAGGCTGGCCTAATAGTGCATACGGGGTGTCTAAGAT TGGAGTGACAGTCCTGTCCAGAATCCTTGCCAGGAAACTCAATGAGCAGAGGAGAGGG GACAAGATCCTTCTGAATGCCTGCTGCCCTGGCTGGGTCAGAACCGACATGGCAGGAC CACAAGCCACCAAAAGCCCAGAAGAAGGAGCAGAGACCCCTGTGTACTTGGCCCTTTT GCCTCCAGATGCAGAGGGACCTCATGGGCAGTTTGTTCAAGATAAAAAAGTGGAACAA TGGT**GA**ACTCAGCTCTTTGTACAGCTCCCATCTGTAGCCTGTCCTAAAGGGGA
	ORF Start: ATG at 101	ORF Stop: TGA at 932
	SEQ ID NO: 282	277 aa	MW at 30547.7kD
NOV98a, CG59669-01 Protein Sequence	MSSCSRVALVTGANKGIGFAITRDLCRKFSGDVVLTARDEARGRAAVQQLQAEGLSPR FHQLDIDDPQSIRALRDFLRKEYGGLNVLVNNAGIAFRSTDLTHFHILREAAMKTNFF GTQAVCTELLPLIKTQGRVVNISSLISLEALKNCSLELQQKFRSETITEEELVGLMNK FVEDTKKGVHAKEGWPNSAYGVSKIGVTVLSRILARKLNEQRRGDKILLNACCPGWVR TDMAGPQATKSPEEGAETPVYLALLPPDAEGPHGQFVQDKKVEQW	

Further analysis of the NOV98a protein yielded the following properties shown in Table 98B.

<table>
<tr><td colspan="2">Table 98B. Protein Sequence Properties NOV98a</td></tr>
</table>

PSort analysis:	0.4766 probability located in mitochondrial matrix space; 0.4500 probability located in cytoplasm; 0.1822 probability located in mitochondrial inner membrane; 0.1822 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV98a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 98C.

<table>
<tr><td colspan="5">Table 98C. Geneseq Results for NOV98a</td></tr>
<tr>
<td>Geneseq Identifier</td>
<td>Protein/Organism/Length [Patent #, Date]</td>
<td>NOV98a Residues/ Match Residues</td>
<td>Identities/ Similarities for the Matched Region</td>
<td>Expect Value</td>
</tr>
<tr>
<td>AAW51011</td>
<td>Human liver carbonyl reductase - Homo sapiens, 277 aa. [US5756299-A, 26-MAY-1998]</td>
<td>1..277
1..277</td>
<td>236/277 (85%)
252/277 (90%)</td>
<td>e-134</td>
</tr>
</table>

AAU33100	Novel human secreted protein #3591 - Homo sapiens, 175 aa. [WO200179449-A2, 25-OCT-2001]	142..277 39..174	119/136 (87%) 128/136 (93%)	2e-66
AAM73641	Human bone marrow expressed probe encoded protein SEQ ID NO: 33947 - Homo sapiens, 123 aa. [WO200157276-A2, 09-AUG-2001]	1..97 1..97	86/97 (88%) 92/97 (94%)	7e-43
AAM60948	Human brain expressed single exon probe encoded protein SEQ ID NO: 33053 - Homo sapiens, 123 aa. [WO200157275-A2, 09-AUG-2001]	1..97 1..97	86/97 (88%) 92/97 (94%)	7e-43
AAM33832	Peptide #7869 encoded by probe for measuring placental gene expression - Homo sapiens, 123 aa. [WO200157272-A2, 09-AUG-2001]	1..97 1..97	86/97 (88%) 92/97 (94%)	7e-43

In a BLAST search of public sequence databases, the NOV98a protein was found to have homology to the proteins shown in the BLASTP data in Table 98D.

Table 98D. Public BLASTP Results for NOV98a				
Protein Accession Number	Protein/Organism/Length	NOV98a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q924V2	CARBONYL REDUCTASE 2 - Cricetulus griseus (Chinese hamster), 277 aa.	1..277 1..277	243/277 (87%) 260/277 (93%)	e-139
Q91X28	SIMILAR TO CARBONYL REDUCTASE 1 - Mus musculus (Mouse), 277 aa.	1..277 1..277	244/277 (88%) 256/277 (92%)	e-139
Q924V3	CARBONYL REDUCTASE 1 - Cricetulus griseus (Chinese hamster), 277 aa.	1..277 1..277	241/277 (87%) 256/277 (92%)	e-137
P48758	Carbonyl reductase [NADPH] 1 (EC 1.1.1.184) (NADPH-dependent carbonyl reductase 1) - Mus musculus (Mouse), 276 aa.	2..277 1..276	240/276 (86%) 253/276 (90%)	e-136
JC5284	carbonyl reductase (NADPH) (EC 1.1.1.184), inducible - rat, 277 aa.	1..277 1..277	236/277 (85%) 249/277 (89%)	e-134

PFam analysis predicts that the NOV98a protein contains the domains shown in the Table 98E.

5

Table 98E. Domain Analysis of NOV98a			
Pfam Domain	**NOV98a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
adh_short: domain 1 of 1	4..274	67/286 (23%) 185/286 (65%)	1.6e-38

Example 99.

The NOV99 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 99A.

Table 99A. NOV99 Sequence Analysis		
	SEQ ID NO: 283	1001 bp
NOV99a, CG58624-01 DNA Sequence	CTTGGTATAAGTAAGTGCTCGTCAATGTTGGCTACTCTCAATGTCAGAGCCGCAGCCG CGGGGCGCAGAGCGCGATCTCTACCGGGACACGTGGGTGCGATACCTGGGCTATGCCA ATGAGGTGGGCGAGGCTTTCCGCTCTCTTGTGCCAGCGGCGGTGGTGTGGCTGAGCTA TGGCGTGGCCAGCTCCTACGTGCTGGCGGATGCCATTGACAAAGGCAAGAAGGCTGGA GAGGTGCCCAGCCCTGAAGCAGGCCGCAGCGCCAGGGTGACTGTGGCTGTGGTGGACA CCTTTGTATGGCAGGCTCTAGCCTCTGTGGCCATTCCGGGCTTCACCATCAACCGCGT GTGTGCTGCCTCTCTCTATGTCCTGGGCACTGCCACCCGCTGGCCCCTGGCTGTCCGC AAGTGGACCACCACCGCGCTTGGGCTGTTGACCATCCCCATCATTATCCACCCCATTG ACAGGGATCATCCACTCTCCAGTGATGAGAGTGGATCATCCAGTCTCCAGCACGAAGG GCCAGGGGTCCCACAGGTGAGTGGAGCCCCAGCAGCCCCCTCAGCTCTGCGTGCCCAT GTACTGGTCTTCTCCCTGGCTCTATACTCAGTGTTCAAGGGGTTGGACGGGGCTTGGG CCGCGGAGCTGCGCCTGGCTTTGCTGCTCCACAAGGGCACCGTGGCTGTCAGCCTGTC CCTGCAACTGCTGCAGAGCCACGTAGGGTTACAGGTGGTGGCTGGCTGTGGGATCCAC TTCTTGTGCATGACACTTCTAGGCATCCGGCTGGGTGCGGCTCTGGCACAGTCAGCAG GGCCTCTGCACCAGCTGGCCCAGTCTGTGCTAGAGGGCATGGTGGCTGGCACCTTCCT CTATACCACCTTTCTGGAAATCTTTCCACAGGAGCTGGCGACTTCTGAGCAAAGGATC CTCAAGGTCATTCTGCTCCTAGAAGGGTGTGCCCTGCTCACTGGCCTGCTCTTCATCC ATATCTAGGGGGCTT
	ORF Start: ATG at 41	ORF Stop: TAG at 992
	SEQ ID NO: 284	317 aa	MW at 33737.8kD
NOV99a, CG58624-01 Protein Sequence	MSEPQPRGAERDLYRDTWVRYLGYANEVGEAFRSLVPAAVVWLSYGVASSYVLADAID KGKKAGEVPSPEAGRSARVTVAVVDTFVWQALASVAIPGFTINRVCAASLYVLGTATR WPLAVRKWTTTALGLLTIPIIIHPIDRDHPLSSDESGSSSLQHEGPGVPQVSGAPAAP SALRAHVLVFSLALYSVFKGLDGAWAAELRLALLLHKGTVAVSLSLQLLQSHVGLQVV AGCGIHFLCMTLLGIRLGAALAQSAGPLHQLAQSVLEGMVAGTFLYTTFLEIFPQELA TSEQRILKVILLLEGCALLTGLLFIHI		

Further analysis of the NOV99a protein yielded the following properties shown in

5   Table 99B.

Table 99B. Protein Sequence Properties NOV99a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 55 and 56

392

A search of the NOV99a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 99C.

Table 99C. Geneseq Results for NOV99a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV99a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM93835	Human polypeptide, SEQ ID NO: 3905 - Homo sapiens, 324 aa. [EP1130094-A2, 05-SEP-2001]	140..317 141..324	134/184 (72%) 145/184 (77%)	3e-63
AAY52394	Human transmembrane protein HP10528 - Homo sapiens, 324 aa. [WO9955862-A2, 04-NOV-1999]	140..317 141..324	134/184 (72%) 145/184 (77%)	3e-63
AAY84895	A human proliferation and apoptosis related protein - Homo sapiens, 324 aa. [WO200023589-A2, 27-APR-2000]	140..317 141..324	134/184 (72%) 145/184 (77%)	3e-63
AAB43291	Human ORFX ORF3055 polypeptide sequence SEQ ID NO:6110 - Homo sapiens, 323 aa. [WO200058473-A2, 05-OCT-2000]	140..317 140..323	134/184 (72%) 145/184 (77%)	3e-63
AAM93650	Human polypeptide, SEQ ID NO: 3514 - Homo sapiens, 324 aa. [EP1130094-A2, 05-SEP-2001]	140..317 141..324	133/184 (72%) 144/184 (77%)	2e-62

5   In a BLAST search of public sequence databases, the NOV99a protein was found to have homology to the proteins shown in the BLASTP data in Table 99D.

Table 99D. Public BLASTP Results for NOV99a				
Protein Accession Number	Protein/Organism/Length	NOV99a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UDX5	WUGSC:H_DJ0539M06.2 PROTEIN - Homo sapiens (Human), 166 aa.	1..152 1..152	145/152 (95%) 145/152 (95%)	6e-78
Q9CRB8	2610507A21RIK PROTEIN (1700020C11RIK PROTEIN) - Mus musculus (Mouse), 166 aa.	1..168 1..164	133/168 (79%) 143/168 (84%)	8e-69

393

Q9CZX4	2610507A21RIK PROTEIN - Mus musculus (Mouse), 166 aa.	1..143 1..143	125/143 (87%) 133/143 (92%)	2e-68
Q9NY26	IRT1 PROTEIN (SIMILAR TO ZINC/IRON REGULATED TRANSPORTER-LIKE) (HYPOTHETICAL 34.2 KDA PROTEIN) (UNKNOWN) (PROTEIN FOR MGC:14180) - Homo sapiens (Human), 324 aa.	140..317 141..324	134/184 (72%) 145/184 (77%)	1e-62
Q9Y380	CGI-71 PROTEIN - Homo sapiens (Human), 324 aa.	140..317 141..324	134/184 (72%) 145/184 (77%)	1e-62

PFam analysis predicts that the NOV99a protein contains the domains shown in the Table 99E.

Table 99E. Domain Analysis of NOV99a			
Pfam Domain	**NOV99a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Syndecan: domain 1 of 1	235..255	9/21 (43%) 16/21 (76%)	6.9
Zip: domain 1 of 1	174..313	52/178 (29%) 108/178 (61%)	2.3e-15

Example 100.

The NOV100 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 100A.

Table 100A. NOV100 Sequence Analysis		
	SEQ ID NO: 285	987 bp
NOV100a, CG59679-01 DNA Sequence	AGACGCTCACACAGACAACCTCAAGTCCAGCAACATCTTAGTAGCCCAAAATCGACTG CTTTAGTTCTTCTGGTGGGTGCCTCTCACTGTCCACTCGGCTATGCCATCCTGCAGTC GCATTGCACTGGTGACTGGAGCTAATAAGGGCATTGGCTTTGCGATCACTCGTGACCT GTGTCAGCAATTCTCAGGGGATGTGGTGCTCACTGCACGGGACGAGGCACGGGGCCTT GCGGCAGTGCAGAAGCTGCAGGCTGAGGGCCTGATTCCTCGCTTCCACCAGCTGGACA TCAATGACCCTCAGAGCATCCATGCACTTCGCAACTTTCTGCTCAAGGAGTACGGAGG CCTGGATGTGTGGTCAACAACGCGGGCATTGGCGTGCTTTTCAAAGTGGATGACCCA ACACCCTTCGACATTCAAGCTGAGGTGACACTGAAGACGAACTTTTTTGCCACTAGAA ATGTCTGCACTGAGTTACTGCCTATAATGAAACCACATGGTAGAGTGGTGAACATCAG CAGTCTGCAGGGGTTAAAAGCCCTTGAGAACTGCAGGGAAGATCTTCAGGAAAAGTTC CGATGTGACACACTTACCGAGGTGGACCTGGTCGACCTCATGAAAAAGTTTGTGGAGG ATACAAAAAATGAAGTCCATGAGAGGGAAGGTTGGCCAGACTCGGCTTACGGGGTGTC GAAGCTGGGGGTGACAGTCCTTACGAGGATCCTGGCCCGGCAGCTGGATGAAAAGAGG AAAGCGGACAGGATTCTGCTCAATGCCTGCTGCCCGGGATGGGTGAAGACCGACATGG CGAGGGACCAGGGCTCCCGGACCGTGGAAGAGGGGGCCGAAACCCCCGTTTACTTGGC TCTCCTGCCTCCAGATGCCACTGAACCTCACGGCCAGCTAGTCCGTGACAAAGTTGTG CAAACTTGGTGAACGTCTGCTCTGGGGCTTAATTGTTTGATAAACGTTAGCGGGAGAG A

394

	ORF Start: ATG at 101	ORF Stop: TGA at 938	
	SEQ ID NO: 286	279 aa	MW at 31007.2kD
NOV100a, CG59679-01 Protein Sequence	MPSCSRIALVTGANKGIGFAITRDLCQQFSGDVVLTARDEARGLAAVQKLQAEGLIPR FHQLDINDPQSIHALRNFLLKEYGGLDVLVNNAGIGVLFKVDDPTPFDIQAEVTLKTN FFATRNVCTELLPIMKPHGRVVNISSLQGLKALENCREDLQEKFRCDTLTEVDLVDLM KKFVEDTKNEVHEREGWPDSAYGVSKLGVTVLTRILARQLDEKRKADRILLNACCPGW VKTDMARDQGSRTVEEGAETPVYLALLPPDATEPHGQLVRDKVVQTW		

Further analysis of the NOV100a protein yielded the following properties shown in Table 100B.

Table 100B. Protein Sequence Properties NOV100a	
PSort analysis:	0.3600 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1808 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV100a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5   several homologous proteins shown in Table 100C.

Table 100C. Geneseq Results for NOV100a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV100a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW51011	Human liver carbonyl reductase - Homo sapiens, 277 aa. [US5756299-A, 26-MAY-1998]	1..279 1..277	198/279 (70%) 233/279 (82%)	e-112
AAU33100	Novel human secreted protein #3591 - Homo sapiens, 175 aa. [WO200179449-A2, 25-OCT-2001]	145..279 40..174	88/135 (65%) 110/135 (81%)	2e-48
AAG46601	Arabidopsis thaliana protein fragment SEQ ID NO: 58644 - Arabidopsis thaliana, 302 aa. [EP1033405-A2, 06-SEP-2000]	3..259 20..283	106/268 (39%) 157/268 (58%)	6e-43
AAG46600	Arabidopsis thaliana protein fragment SEQ ID NO: 58643 - Arabidopsis thaliana, 316 aa. [EP1033405-A2, 06-SEP-2000]	3..259 34..297	106/268 (39%) 157/268 (58%)	6e-43

AAG46599	Arabidopsis thaliana protein fragment SEQ ID NO: 58642 - Arabidopsis thaliana, 327 aa. [EP1033405-A2, 06-SEP-2000]	3..259 45..308	106/268 (39%) 157/268 (58%)	6e-43

In a BLAST search of public sequence databases, the NOV100a protein was found to have homology to the proteins shown in the BLASTP data in Table 100D.

Table 100D. Public BLASTP Results for NOV100a				
Protein Accession Number	Protein/Organism/Length	NOV100a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9JJN7	CARBONYL REDUCTASE (EC 1.1.1.184) (CARBONYL REDUCTASE 3) - Cricetulus griseus (Chinese hamster), 277 aa.	1..279 1..277	246/279 (88%) 262/279 (93%)	e-140
AAH02812	CARBONYL REDUCTASE 3 - Homo sapiens (Human), 277 aa.	1..279 1..277	227/279 (81%) 246/279 (87%)	e-126
O75828	Carbonyl reductase [NADPH] 3 (EC 1.1.1.184) (NADPH-dependent carbonyl reductase 3) - Homo sapiens (Human), 276 aa.	3..279 2..276	226/277 (81%) 245/277 (87%)	e-126
Q924V2	CARBONYL REDUCTASE 2 - Cricetulus griseus (Chinese hamster), 277 aa.	1..279 1..277	206/279 (73%) 244/279 (86%)	e-119
Q91X28	SIMILAR TO CARBONYL REDUCTASE 1 - Mus musculus (Mouse), 277 aa.	1..279 1..277	204/279 (73%) 240/279 (85%)	e-116

PFam analysis predicts that the NOV100a protein contains the domains shown in the Table 100E.

Table 100E. Domain Analysis of NOV100a			
Pfam Domain	NOV100a Match Region	Identities/ Similarities for the Matched Region	Expect Value
adh_short: domain 1 of 1	4..277	77/316 (24%) 186/316 (59%)	5.2e-31

Example 101.

The NOV101 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 101A.

Table 101A. NOV101 Sequence Analysis

	SEQ ID NO: 287	1011 bp	
NOV101a, CG59644-01 DNA Sequence	CTCCTCGGGGGGGCGGCGGCGGCGATGTTCTCGGTCCTCTCGTACGGGCGGCTGGTGG CCCGCGCCGTGCTCGGCGGCCTCTCGCAGACCGACCCCAGGGCCGGCGGCGGCGGCGG CGGCGCCGTGCTCGGCGGCCTCTCGCAGACCGACCCCAGGGCCGGCGGCGGCGGCGGC GGCGACTACGGACTGGTGACGGCCGGCTGCGGCTTCGGGAAGGACTTCCGTAAGGGCC TCCTCAAGAAGGGCGCGTGCTACGGGGACGACGCGTGCTTCGTGGCCCGGCACCGTTC CGCGGACGTGCTCGGTGTTGCAGATGGTGTAGGAGGCTGGAGAGACTATGGAGTTGAT CCATCTCAATTCTCAGGGACTTTAATGCGGACGTGTGAACGTTTAGTAAAAGAAGGAC GGTTCGTACCTAGTAATCCCATTGGAATTCTCACCACAAGCTACTGTGAGTTGCTGCA AAATAAAGTCCCTTTGCTCGGTAGCAGCACCGCCTGCATTGTGGTGCTGGACAGAACC AGCCACCGCTTACACACAGCAAACCTGGGCGATTCAGGCTTCCTGGTTGTCAGGGGTG GTGAAGTCGTGCACCGATCAGATGAGCAGCAGCATTACTTCAACACTCCATTCCAGCT CTCAATCGCTCCCCCTGAAGCCGAGGGAGTCGTCTTGAGCGACAGTCCGGATGCTGCT GATAGCACGTCTTTCGATGTCCAGCTAGGAGACATTATCCTGACGGCAACAGATGGAC TCTTTGACAACATGCCTGATTATATGATTCTTCAGGAGCTAAAAAAGTTAAAGAATTC AAATTATGAGAGTATACAACAGACTGCCAGAAGCATTGCTGAGCAAGCTCATGAGCTG GCCTATGACCCAAATTATATGTCACCTTTTGCACAGTTTGCATGTGACAATGGATTGA ATGTGAGAGGTGGTGGAAAGCCAGATGACATCACCGTCCTTCTTTCAATAGTGGCTGA GTATACAGACTAGCTGAGGTGTCAA	
	ORF Start: ATG at 25	ORF Stop: TAG at 997	
	SEQ ID NO: 288	324 aa	MW at 34311.1kD
NOV101a, CG59644-01 Protein Sequence	MFSVLSYGRLVARAVLGGLSQTDPRAGGGGGAVLGGLSQTDPRAGGGGGDYGLVTA GCGFGKDFRKGLLKKGACYGDDACFVARHRSADVLGVADGVGGWRDYGVDPSQFSGTL MRTCERLVKEGRFVPSNPIGILTTSYCELLQNKVPLLGSSTACIVVLDRTSHRLHTAN LGDSGFLVVRGGEVVHRSDEQQHYFNTPFQLSIAPPEAEGVVLSDSPDAADSTSFDVQ LGDIILTATDGLFDNMPDYMILQELKKLKNSNYESIQQTARSIAEQAHELAYDPNYMS PFAQFACDNGLNVRGGGKPDDITVLLSIVAEYTD		

Further analysis of the NOV101a protein yielded the following properties shown in
5   Table 101B.

Table 101B. Protein Sequence Properties NOV101a

PSort analysis:	0.5708 probability located in mitochondrial matrix space; 0.4996 probability located in mitochondrial intermembrane space; 0.2852 probability located in mitochondrial inner membrane; 0.2852 probability located in mitochondrial outer membrane
SignalP analysis:	Likely cleavage site between residues 23 and 24

A search of the NOV101a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 101C.

Table 101C. Geneseq Results for NOV101a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV101a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB85357	Human phosphatase (PP) (clone ID 3402521CD1) - Homo sapiens, 304 aa. [WO200153469-A2, 26-JUL-2001]	1..324 1..304	304/324 (93%) 304/324 (93%)	e-173
AAU32112	Novel human secreted protein #2603 - Homo sapiens, 304 aa. [WO200179449-A2, 25-OCT-2001]	25..324 6..304	272/300 (90%) 274/300 (90%)	e-156
AAG52267	Arabidopsis thaliana protein fragment SEQ ID NO: 66421 - Arabidopsis thaliana, 348 aa. [EP1033405-A2, 06-SEP-2000]	71..320 99..340	101/261 (38%) 133/261 (50%)	4e-33
AAG52266	Arabidopsis thaliana protein fragment SEQ ID NO: 66420 - Arabidopsis thaliana, 374 aa. [EP1033405-A2, 06-SEP-2000]	71..320 125..366	101/261 (38%) 133/261 (50%)	4e-33
AAG52265	Arabidopsis thaliana protein fragment SEQ ID NO: 66419 - Arabidopsis thaliana, 467 aa. [EP1033405-A2, 06-SEP-2000]	71..320 218..459	101/261 (38%) 133/261 (50%)	4e-33

In a BLAST search of public sequence databases, the NOV101a protein was found to have homology to the proteins shown in the BLASTP data in Table 101D.

Table 101D. Public BLASTP Results for NOV101a				
Protein Accession Number	**Protein/Organism/Length**	**NOV101a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9W0E2	CG12091 PROTEIN - Drosophila melanogaster (Fruit fly), 321 aa.	1..320 1..320	163/322 (50%) 218/322 (67%)	1e-83
Q9W3R1	CG15035 PROTEIN - Drosophila melanogaster (Fruit fly), 374 aa.	55..319 109..373	127/266 (47%) 178/266 (66%)	1e-64
O18183	W09D10.4 PROTEIN - Caenorhabditis elegans, 330 aa.	4..320 7..330	136/331 (41%) 198/331 (59%)	2e-60

| Q9VAH4 | CG7615 PROTEIN - Drosophila melanogaster (Fruit fly), 314 aa. | 35..319 26..309 | 122/285 (42%) 168/285 (58%) | 2e-56 |
| Q9SUK9 | HYPOTHETICAL 36.2 KDA PROTEIN - Arabidopsis thaliana (Mouse-ear cress), 335 aa. | 71..320 86..327 | 101/261 (38%) 133/261 (50%) | 1e-32 |

PFam analysis predicts that the NOV101a protein contains the domains shown in the Table 101E.

<table>
<tr><th colspan="4">Table 101E. Domain Analysis of NOV101a</th></tr>
<tr><th>Pfam Domain</th><th>NOV101a Match Region</th><th>Identities/ Similarities for the Matched Region</th><th>Expect Value</th></tr>
<tr><td>PP2C: domain 1 of 1</td><td>147..191</td><td>13/48 (27%) 36/48 (75%)</td><td>0.26</td></tr>
</table>

Example 102.

The NOV102 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 102A.

<table>
<tr><th colspan="4">Table 102A. NOV102 Sequence Analysis</th></tr>
<tr><td></td><td>SEQ ID NO: 289</td><td colspan="2">523 bp</td></tr>
<tr><td>NOV102a, CG59662-01 DNA Sequence</td><td colspan="3">AGTCCCAGTACTATCAGCCATGGTCAACCACACCATGTTCTTCGACGTTGCTGTCGAC AGTGAGCCCTTGGACCACGTCTCCTTTGAGCTGTTTGCAGAAAAGTTTCCAAAGACAG CAGAAAACGTTCGTGCTCTGAGCACTGAAGAGAAAGGATTTGGTTATAAGGGTCCCTG CTTTCACAGAATTATACCAGCATTTATGTGTCAGGGTGGTGACTTCACGCACCATAAT GGCACTGGTGGCAAGTCCATCTACGGGGAGAAATTTGAAGATGAGAAATTTATCCTAA AGCGTACAGGTCCTGGCATCTTGTCCATGGCAAATTCTGGACCCAACACAAACTGTTC CGTTTTTTTCATCTGCACTGCCAAGACGGGGTGGTTGGATGGCAAGCATGTAGTCTTT GGCAAGGTGAAAGAAGGCATGAATATTTTGGAGGCCATAGAGCAATTTGGGTCCAGGA ATGGCAAGACCAGCAAGAAGACCACCATTGCTGACTGTGGACAGCTCTGG<u>TAA</u>GTTTG A</td></tr>
<tr><td></td><td>ORF Start: ATG at 20</td><td colspan="2">ORF Stop: TAA at 515</td></tr>
<tr><td></td><td>SEQ ID NO: 290</td><td>165 aa</td><td>MW at 18237.7kD</td></tr>
<tr><td>NOV102a, CG59662-01 Protein Sequence</td><td colspan="3">MVNHTMFFDVAVDSEPLDHVSFELFAEKFPKTAENVRALSTEEKGFGYKGPCFHRIIP AFMCQGGDFTHHNGTGGKSIYGEKFEDEKFILKRTGPGILSMANSGPNTNCSVFFICT AKTGWLDGKHVVFGKVKEGMNILEAIEQFGSRNGKTSKKTTIADCGQLW</td></tr>
</table>

Further analysis of the NOV102a protein yielded the following properties shown in Table 102B.

<table>
<tr><th colspan="2">Table 102B. Protein Sequence Properties NOV102a</th></tr>
<tr><td>PSort analysis:</td><td>0.6400 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)</td></tr>
</table>

399

SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV102a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 102C.

Table 102C. Geneseq Results for NOV102a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV102a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU01195	Human cyclophilin A protein - Homo sapiens, 165 aa. [WO200132876-A2, 10-MAY-2001]	1..164 1..164	141/164 (85%) 148/164 (89%)	1e-80
AAW56028	Calcineurin protein - Mammalia, 165 aa. [WO9808956-A2, 05-MAR-1998]	1..164 1..164	141/164 (85%) 148/164 (89%)	1e-80
AAG65275	Haematopoietic stem cell proliferation agent related human protein #2 - Homo sapiens, 164 aa. [JP2001163798-A, 19-JUN-2001]	2..164 1..163	140/163 (85%) 147/163 (89%)	5e-80
AAP90431	Cyclophilin - Homo sapiens (human), 164 aa. [EP326067-A, 02-AUG-1989]	2..164 1..163	140/163 (85%) 147/163 (89%)	5e-80
AAG03831	Human secreted protein, SEQ ID NO: 7912 - Homo sapiens, 165 aa. [EP1033401-A2, 06-SEP-2000]	1..164 1..164	140/164 (85%) 147/164 (89%)	8e-80

In a BLAST search of public sequence databases, the NOV102a protein was found to have homology to the proteins shown in the BLASTP data in Table 102D.

Table 102D. Public BLASTP Results for NOV102a				
Protein Accession Number	Protein/Organism/Length	NOV102a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC39529	SEQUENCE 26 FROM PATENT	1..164 1..164	141/164 (85%) 148/164 (89%)	4e-80

	(Human), 165 aa.			
Q9BRU4	PEPTIDYLPROLYL ISOMERASE A (CYCLOPHILIN A) - Homo sapiens (Human), 165 aa.	1..164 1..164	140/164 (85%) 147/164 (89%)	2e-79
P05092	Peptidyl-prolyl cis-trans isomerase A (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin A) (Cyclosporin A-binding protein) - Homo sapiens (Human),, 164 aa.	2..164 1..163	140/163 (85%) 147/163 (89%)	2e-79
Q96IX3	PEPTIDYLPROLYL ISOMERASE A (CYCLOPHILIN A) - Homo sapiens (Human), 165 aa.	1..164 1..164	140/164 (85%) 147/164 (89%)	5e-79
P04374	Peptidyl-prolyl cis-trans isomerase A (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin A) (Cyclosporin A-binding protein) - Bos taurus (Bovine), and, 163 aa.	2..164 1..163	138/163 (84%) 147/163 (89%)	7e-79

PFam analysis predicts that the NOV102a protein contains the domains shown in the Table 102E.

Table 102E. Domain Analysis of NOV102a			
Pfam Domain	NOV102a Match Region	Identities/ Similarities for the Matched Region	Expect Value
pro_isomerase: domain 1 of 1	5..165	105/180 (58%) 141/180 (78%)	4.2e-91

Example 103.

The NOV103 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 103A.

Table 103A. NOV103 Sequence Analysis		
	SEQ ID NO: 291	8860 bp
NOV103a, CG59773-01 DNA Sequence	GGATCCTTGAGGGCACTGGTGCGACTTTCAGGTGAGGTCTTAGCAGATGAAAGCGGCT GGCTGTGGCCCGCGCCAGTAGTGCTTTCTGCTCCGCACTCGCCGTGAGCCAGGTGTGC AACCGGATTTGGGGCGAGGGTCGCGCTGGCTACCTCGCATGCGCAGAGCCGGAAGCCC GCTGACCGGACTACAGCTCCCAGAAGAGCCTTGTGGAGGCCGCAGACGCGAAGCCGCT GGCGCCATCTTGAAATCTGATCCTCCATCCCCGAGGCTTTGCGTCTGCGCGGCCGGCC GCTGCTGCTCCGGGAGCCCAGTCTGCTAAAAGGGGAGGACGTTGAGGACGCGGCGGCT GGCGGGAGAGACAGCTGGGGAGAGACATGGCAGGGTCGGAGCGCGGCCTGCGCCTCTG TCACTCAGCATCCTCTTAGCGTTTCCACGCCCGCCCCCTGCCCGAGGGGCGGGGCTG ACGGCTCTGGTACCCGGAGTCGGCGCGCGGGGCAGGGGCGCGCCCCTGCAGAGTGGGG ACCCCACTGGGCTGTGCCATGCTGACCGGAGACCACCGAGGCGGGAGACAGAGCGCGG CGAAGAGCCATTGAGTGGTCACCCAGTAGCCGCCGCCGCCGCCGCCTCGGGAAGCTTG CCACCCGCTAGGAGGGAAGATGAAGGAGATTTGCAGGATCTGTGCCCGAGAGCTGTGT GGAAACCAGCGGCGCTGGATCTTCCACACGGCGTCCAAGCTCAATCTCCAGGTTCTGC	

401

```
TTTCGCACGTCTTGGGCAAGGATGTCCCCCGCGATGGCAAAGCCGAGTTCGCTTGCAG
CAAGTGTGCTTTCATGCTTGATCGAATCTATCGATTCGACACAGTTATTGCCCGGATT
GAAGCGCTTTCTATTGAGCGCTTGCAAAAGCTGCTACTGGAGAAGGATCGCCTCAAGT
TCTGCATTGCCAGTATGTATCGGAAGAATAACGATGACTCTGGCGCGGAGATCAAGGC
GGGGAATGGGACGGTTGACATGTCCGTCTTACCCGATGCGAGATACTCTGCACTGCTC
CAGGAGGACTTCGCCTATTCAGGGTTTGAGTGCTGGGTGGAGAATGAGGATCAGATCC
AGGAGCCACACAGCTGCCATGGTTCAGAAGGCCCTGGAAACCGACCCAGGAGATGCCG
TGGTTGTGCCGCTTTGCGGGTTGCTGATTCTGACTATGAAGCCATTTGTAAGGTACCT
CGAAAGGTGGCCAGAAGTATCTCCTGCGGCCCTTCTAGCAGGTGGTCGACCAGCATTT
GCACTGAAGAACCAGCGTTGTCTGAGGTTGGGCCACCCGACTTAGCAAGCACAAAGGT
ACCCCCAGATGGAGAAAGCATGGAGGAAGAGACGCCTGGTTCCTCTGTGGAATCTTTG
GATGCAAGCGTCCAGGCTAGCCCTCCACAACAGAAAGATGAGGAGACTGAGAGAAGTG
CAAAGGAACTTGGAAAGTGTGACTGTTGTTCAGATGATCAGGCTCCGCAGCATGGGTG
TAATCACAAGCTGGAATTAGCTCTTAGCATGATTAAAGGTCTTGATTATAAGCCCATC
CAGAGCCCCCGAGGGAGCAGGCTTCCGATTCCAGTGAAATCCAGCCTACCTGGAGCCA
AGCCTGGCCCTAGCATGACAGATGGAGTTAGTTCCGGTTTCCTTAACAGGTCTTTGAA
ACCCCTTTACAAGACACCTGTGAGTTATCCCTTGGAGCTTTCAGACCTGCAGGAGCTG
TGGGATGATCTCTGTGAAGATTATTTGCCGCTCCGGGTCCAGCCCATGACTGAAGAGT
TGCTGAAACAACAAAAGCTGAATTCACATGAGACCACTATAACTCAGCAGTCTGTATC
TGATTCCCACTTGGCAGAACTCCAGGAAAAAATCCAGCAAACAGAGGCCACCAACAAG
ATTCTTCAAGAGAAACTTAATGAAATGAGCTATGAACTAAAGTGTGCTCAGGAGTCGT
CTCAAAAGCAAGATGGTACAATTCAGAACCTCAAGGAAACTCTGAAAAGCAGGGAACG
TGAGACTGAGGAGTTGTACCAGGTAATTGAAGGTCAAAATGACACAATGGCAAAGCTT
CGAGAAATGCTGCACCAAAGCCAGCTTGGACAACTTCACAGCTCAGAGGGTACTTCTC
CAGCTCAGCAACAGGTAGCTCTGCTTGATCTTCAGAGTGCTTTATTCTGCAGCCAACT
TGAAATACAGAAGCTCCAGAGGGTGGTACGACAGAAAGAGCGCCAACTGGCTGATGCC
AAACAATGTGTGCAATTTGTAGAGGCTGCAGCACACGAGAGTGAACAGCAGAAAGAGG
CTTCTTGGAAACATAACCAGGAATTGCGAAAAGCCTTGCAGCAGCTACAAGAAGAATT
GCAGAATAAGAGCCAACAGCTTCGTGCCTGGGAGGCTGAAAAATACAATGAGATTCGA
ACCCAGGAACAAAACATCCAGCACCTAAACCATAGTCTGAGTCACAAGGAGCAGTTGC
TTCAGGAATTTCGGGAGCTCCTACAGTATCGAGATAACTCAGACAAAACCCTTGAAGC
AAATGAAATGTTGCTTGAGAAACTTCGCCAGCGAATACATGATAAAGCTGTTGCTCTG
GAGCGGGCTATAGATGAAAAATTCTCTGCTCTAGAAGAGAAAGAAAAAGAACTGCGCC
AGCTTCGTCTTGCTGTGAGAGAGCGAGATCATGACTTAGAGAGACTGCGCGATGTCCT
CTCCTCCAATGAAGCTACTATGCAAAGTATGGAGAGTCTCCTGAGGGCCAAAGGCCTG
GAAGTGGAACAGTTATCTACTACCTGTCAAAACCTCCAGTGGCTGAAAGAAGAAATGG
AAACCAAATTTAGCCGTTGGCAGAAGGAACAAGAGAGTATCATTCAGCAGTTACAGAC
GTCTCTTCATGATAGGAACAAAGAAGTGGAGGATCTTAGTGCAACACTGCTCTGCAAA
CTTGGACCAGGGCAGAGTGAGATAGCAGAGGAGCTGTGCCAGCGTCTACAGCGAAAGG
AAAGGATGCTGCAGGACCTTCTAAGTGATCGAAATAAACAAGTGCTGGAACATGAAAT
GGAGATTCAAGGCCTGCTTCAGTCTGTGAGCACCAGGGAGCAGGAAAGCCAAGCTGCT
GCAGAGAAGTTGGTGCAAGCCTTAATGGAAAGAAATTCAGAATTACAGGCCCTGCGCC
AATATTTAGGAGGGAGAGACTCCCTGATGTCCCAAGCACCCATCTCTAACCAACAAGC
TGAAGTTACCCCCACTGGCCGTCTTGGAAAACAGACTGATCAAGGTTCAATGCAGATA
CCTTCCAGAGATGATAGCACTTCATTGACTGCCAAAGAGGATGTCAGCATACCCAGAT
CCACATTAGGAGACTTGGACACAGTTGCAGGGCTGGAAAAAGAACTGAGTAATGCCAA
AGAGGAACTTGAACTCATGGCTAAAAAAGAAAGAGAAAGTCAGATGGAACTTTCTGCT
CTACAGTCCATGATGGCTGTGCAGGAAGAAGAGCTGCAGGTGCAGGCTGCTGATATGG
AGTCTCTGACCAGGAACATACAGATTAAAGAAGATCTCATAAAGGACCTGCAAATGCA
ACTGGTTGATCCTGAAGACATACCAGCTATGGAACGCCTGACCCAGGAAGTCTTACTT
CTTCGGGAAAAAGTTGCTTCAGTAGAATCCCAGGGTCAAGAAATTTCAGGAAACCGAA
GACAACAGTTGCTGCTGATGCTAGAAGGACTAGTAGATGAACGGAGTCGGCTCAATGA
GGCCTTACAAGCAGAGAGACAGCTCTATAGCAGTCTGGTGAAGTTCCATGCCCATCCA
GAGAGCTCTGAGAGAGACCGAACTCTGCAGGTGGAACTGGAAGGGGCTCAGGTGTTAC
GCAGTCGGCTAGAAGAAGTTCTTGGAAGAAGCTTGGAGCGCTTAAACAGGCTGGAGAC
CCTGGCCGCCATTGGAGGTGCAGCTGCAGGGGATGACACCGAAGATACAAGCACTGAG
TTCACTGACAGTATTGAGGAGGAGGCTGCACACCATAGTCACCAGCAACTTGTCAAGG
TGGCTTTGGAGAAAAGTCTGGCAACTGTGGAGACCCAGAACCCATCTTTTTCCCCTCC
TTCTCCGATGGGAGGGGACAGTAACAGGTGTCTTCAGGAAGAAATGCTCCACCTGAGG
GCTGAGTTCCACCAGCACTTAGAAGAGAAGAGGAAAGCTGAGGAGGAACTGAAGGAGC
TAAAGGCTCAAATTGAGGAAGCAGGATTCTCCTCAGTGTCCCACATCAGGAACACCAT
GCTGAGCCTTTGCCTTGAGAATGCGGAGCTGAAAGAGCAGATGGGAGAAGCAATGTCT
GATGGATGGGAGATCGAGGAAGACAAGGAGAAGGGCGAGGTGATGGTTGAGACTGTGG
TAACCAAAGAGGGTCTGAGTGAGAGTAGCCTTCAGGCTGAGTTCAGAAAGCTCCAGGG
AAAACTGAAGAATGCCCACAATATCATCAACCTCCTCAAAGAACAACTTGTGCTGAGT
AGCAAGGAAGGGAATAGTAAACTTACTCCAGAGCTCCTTGTGCATCTGACCAGCACCA
TTGAAAGAATAAACACAGAACTGGTTGGTTCCCCTGGGAAGCACCAACACCAAGAGGA
GGGGAATGTGACTGTGAGGCCTTTCCCCAGACCCCAGAGCCTTGACCTTGGGGCTACC
TTCACAGTGGATGCCCACCAATTGGATAACCAGTCCCAGCCTCGTGACCCTGGGCCTC
AGTCAGCGTTTAGCCTACCAGGGTCCACCCAGCACCTGCGCTCCCAGCTGTCACAATG
CAAACAACGCTATCAAGATCTCCAGGAGAAGCTGCTGCTATCAGAAGCCACTGTCTTT
GCTCAGGCTAACGAGCTGGAGAAATACAGAGTTATGCTTACAGGTGAATCCTTGGTGA
AGCAGGACAGCAAGCAGATCCAGGTGGACCTCCAGGACCTGGGCTATGAGACTTGTGG
CCGAAGCGAGAATGAGGCTGAACGGGAGGGAAACCACCAGTCCTGAGTGTGAGGAGCAC
AACAGCCTCAAGGAAATGGTCCTGATGGAGGGGCTGTGTCTCTGAGCAGGGACGCCGGG
GCTCAACACTGGCTAGTTCCTCTGAGAGGAAGCCCTTGGAGAACCAGCTAGGGAAGCA
GGAAGAGTTCCGGGTATATGGAAAGTCAGAAAACATCTTGGTCCTACGAAAGGACATC
```

402

	AAAGATCTGAAGGCCCAGCTGCAGAATGCCAACAAGGTCATTCAAAACCTCAAGAGCC GGGTCCGGTCCCTCTCAGTTACAAGTGATTATTCGTCTAGTCTGGAAAGACCCCGGAA GCTGAGAGCTGTTGGCACCTTGGAGGGGTCTTCACCTCATAGTGTCCCTGATGAGGAT GAGGGGTGGCTGTCTGATGGCACTGGGGCTTTCTACTCTCCAGGGCTTCAGGCCAAAA AGGACCTGGAGAGTCTCATCCAGAGAGTATCCCAGCTGGAGGCCCAGCTCCCAAAAAA TGGACTAGAAGAGAAGCTGGCTGAGGAGCTGAGATCAGCCTCGTGGCCTGGGAAATAT GATTCCCTGATTCAGGATCAGGCCCGGGAACTGTCTTACCTACGGCAAAAAATACGAG AAGGGAGAGGTATTTGTTATCTTATCACCCGGCATGCAAAAGATACAGTAAAATCTTT TGAGGATCTCCTAAGGAGCAATGACATTGACTACTACCTGGGACAGAGCTTCCGGGAG CAACTCGCCCAGGGAAGCCAGCTGACAGAGAGGCTCACCAGCAAACTCAGCACCAAGG ATCATAAAAGTGAGAAAGATCAAGCTGGACTTGAGCCACTGGCCCTCAGGCTCAGCAG GGAGCTGCAGGAGAAGGAGAAAGTGATTGAAGTCCTGCAGGCCAAGCTGGATGCTCGG TCCCTCACACCCTCCAGCAGCCATGCCTTGTCTGACTCCCACCGCTCTCCCAGCAGCA CCTCTTTCCTGTCTGATGAACTGGAAGCCTGCTCTGACATGGACATAGTCAGCGAGTA CACACACTATGAAGAGAAGAAAGCTTCTCCCAGTCACTCAGATTCCATCCATCATTCG AGTCATTCTGCTGTGTTGTCTTCTAAACCATCATCAACCAGTGCATCTCAGGGGGCTA AGGCCGAATCCAACAGCAACCCCATCAGCTTGCCAACTCCCCAGAATACCCCCAAGGA GGCCAACCAGGCCCATTCAGGCTTTCATTTTCACTCCATACCCAAGCTGGCTAGCCTT CCTCAGGCACCATTGCCCTCAGCTCCATCCAGCTTCCTGCCTTTCAGCCCCACTGGCC CTCTCCTCCTTGGCTGCTGTGTGAGACACCAGTGGTCTCCTTGGCTGAGGCTCAGCAGA GCTACAGATGCTGCAGAAGCAGTTGGGAGAAAGTGCCAGCACTGTTCCTCCTGCTTCC ACAGCTACATTGCTGAGCAACGACTTGGAAGCCGACTCTTCCTACTACCTCAACTCTG CCCAGCCTCACTCTCCTCCAAGGGGCACCATAGAACTGGGAAGAATCCTAGAGCCTGG GTACCTGGGCAGCAGTGGCAAGTGGGATGTGATGAGGCCTCAGAAAGGGAGTGTATCT GGGGACCTATCCTCAGGCTCCTCTGTGTACCAGCTTAACTCCAAACCCACAGGGGCTG ACCTGCTGGAAGAGCATCTTGGTGAAATCCGGAACCTGCGCCAGCGCCTGGAGGAGTC CATCTGCATCAATGACCGCCTACGGGAGCAACTGGAACACCGGCTGACCTCTACTGCT CGTGGAAGGGGATCCACTTCTAACTTCTACAGTCAGGGCCTGGAGTCCATACCTCAGC TCTGCAATGAGAACAGAGTCCTCAGGGAAGACAATCGAAGACTTCAGGCTCAACTGAG TCATGTTTCCAGAGAGCACTCCCAGGAAACAGAAAGCCTGAGGGAGGCTCTGCTGTCC TCTCGATCCCACCTTCAAGAGCTGGAAAAGGAGCTGGAGCACCAGAAGGTGGAAAGGC AGCAGCTTTTGGAAGACTTGAGGGAGAAGCAGCAAGAGGTCTTGCATTTCAGGGAGGA ACGTCTTTCCCTCCAGGAAAACGACTCCAGACTGCAGCACAAGCTGGTTCTCCTGCAG CAACAGTGTGAAGAGAAACAGCAGCTCTTTGAGTCCCTCCAGTCAGAGCTACAAATCT ACGAGGCACTTTATGGCAATTCCAAGAAGGGGCTGAAAGGCTTGGGTTTGGATACTTC TCCAGTAATGAAGACCCCTCCCAAGCTAGAGGGTGATGCTACTGATGGCTCCTTTGCC AATAAGCATGGCCGCCATGTCATTGATCGTCATTGATGACTACAGTGCCCTAAGACAGC AGATTGCGGAGGGCAAGCTGCTGGTCAAAAAGATAGTGTCTCTTGTGAGATCAGCGTG CAGCTTCCCTGGCCTTGAAGCCCAAGGCACAGAGGTGCTAGGCAGCAAAGGTATTCAT GAGCTTCGGAGCAGCACCAGTGCCCTGCACCATGCCCTAGAGGAGTCGGCTTCCCTCC TCACCATGTTCTGGAGAGCAGCAGCCCTGCCAAGCACCCACATCCCTGTGCTGCCTGGCAA AGTGGGAGAATCAACAGAAAGGGAACTTCTGGAACTGAGAACCAAAGTATCCAAACAG GAGCGGCTCCTTCAGAGCACAACTGAGCATCTGAAGAACGCCAACCAGCAGAAGGAGA GCATGGAGCAGTTCATCGTCAGCCAGCTAACCAGAACACATGATGTTTTAAAGAAGGC AAGGACTAACTTAGAGGTGAAATCCCTAAGGGCTCTGCCATGTACTCCAGCCTTG**TGA** CCCTTGCCTTCCAGGAACCATGCAAGAAGCGCAGCCACCAGAAGTCCTTAAAACAGCA GGAAAGGTGGGCCTGTCCCCCTTTTGTGCAGCTACCTATCTGCTGAGGAGCATCTGGG CCTCATTCCTCCAAGTCCACGGGAGGGTCCAGAAGAGGGAGTCAGAGATGTATCCTGG TGGAGCTGGGAGAAAGGCAGAAAGCCTTTCTGACAGCTATGGAATACGATTAGCCAAG GTCCACTTGGCCCAGCACTAAGAAAAAGATGCGTAGTTTGCACAGAAGGTTTTTGTGAT CCTGCCTCTCAACAGCCCCAGCAGCTTGGGAACTAGCAAGAGCACATTTCTTGCCTCA TCAGCTGTCCTGAGATGGAAAACTCAGTGGATATAGGACCCTGATTCCGATGAAAGGG GCACGTGGTCCCAATGCTGGAGCTCCTCTGGCAGGTTCTAAAAGCACACTACTGAGCA GCGGTGCCCTGCCGGACACTGCTGGCGGGGGCTCAGTGAGCACTACTCACAGATCCAC ACCTGACCCTGTTGGGTCGAGTCAGGCTGGGCCTTGGTCTGCACTGTAGCACCTGTGT TCTTTGAGTTCACATCATGAATGTGGTGACTTCCAGATACCATCTCAGGCTTAACCT AGCACATCCTATTTCTTTTCTTCTATGGATATCCAAATTGGACTGACCTCACTTCAAAG TTGCTGTCCCATTTTGTCACCCTATCTTATCTCGGGGAAATTGCAGACTGATGGCCAG ACCAACTCTGTTGAAATTCTTGCATAGAGCAAACCTGTGCTCATTTTTAAGTGGCATG GGAGAGGCCCCCAGCCTAGTAAAGCCTAGTCTGTGTCTTCACAGTGCTGGTAGAATGT GTTTGTGTGTATAAATATATGATATAGATTTATATATGTTGCTAACGCCATATATTGA AGGCCAACATAACTGGTGGACAGGGTGGGTGACAGAAAATGAAAGCCTTTTTGGTGAT TGTTAAAGCAAGATGTGTATAAAGAAATAAATAGTTTTTCTTTC
	ORF Start: ATG at 658	**ORF Stop: TGA at 7828**
	SEQ ID NO: 292	**2390 aa**	**MW at 268843.7kD**
NOV103a, CG59773-01 Protein Sequence	MKEICRICARELCGNQRRWIFHTASKLNLQVLLSHVLGKDVPRDGKAEFACSKCAFML DRIYRFDTVIARIEALSIERLQKLLLEKDRLKFCIASMYRKNNDDSGAEIKAGNGTVD MSVLPDARYSALLQEDFAYSGFECWVENEDQIQEPHSCHGSEGPGNRPRRCRGCAALR VADSDYEAICKVPRKVARSISCGPSSRWSTSICTEEPALSEVGPPDLASTKVPPDGES MEEETPGSSVESLDASVQASPPQQKDEETERSAKELGKCDCCSDDQAPQHGCNHKLEL ALSMIKGLDYKPIQSPRGSRLPIPVKSSLPGAKPGPSMTDGVSSGFLNRSLKPLYKTP VSYPLELSDLQELWDDLCEDYLPLRVQPMTEELLKQQKLNSHETTITQQSVSDSHLAE LQEKIQQTEATNKILQEKLNEMSYELKCAQESSQKQDGTIQNLKETLKSRERETEELY

403

	QVIEGQNDTMAKLREMLHQSQLGQLHSSEGTSPAQQQVALLDLQSALFCSQLEIQKLQ RVVRQKERQLADAKQCVQFVEAAAHESEQQKEASWKHNQELRKALQQLQEELQNKSQQ LRAWEAEKYNEIRTQEQNIQHLNHSLSHKEQLLQEFRELLQYRDNSDKTLEANEMLLE KLRQRIHDKAVALERAIDEKFSALEEKEKELRQLRLAVRERDHDLERLRDVLSSNEAT MQSMESLLRAKGLEVEQLSTTCQNLQWLKEEMETKFSRWQKEQESIIQQLQTSLHDRN KEVEDLSATLLCKLGPGQSEIAEELCQRLQRKERMLQDLLSDRNKQVLEHEMEIQGLL QSVSTREQESQAAAEKLVQALMERNSELQALRQYLGGRDSLMSQAPISNQQAEVTPTG RLGKQTDQGSMQIPSRDDSTSLTAKEDVSIPRSTLGDLDTVAGLEKELSNAKEELELM AKKERESQMELSALQSMMAVQEEELQVQAADMESLTRNIQIKEDLIKDLQMQLVDPED IPAMERLTQEVLLLREKVASVESQGGQEISGNRRQQLLLMLEGLVDERSRLNEALQAER QLYSSLVKFHAHPESSERDRTLQVELEGAQVLRSRLEEVLGRSLERLNRLETLAAIGG AAAGDDTEDTSTEFTDSIEEEAAHHSHQQLVKVALEKSLATVETQNPSFSPPSPMGGD SNRCLQEEMLHLRAEFHQHLEEKRKAEEELKELKAQIEEAGFSSVSHIRNTMLSLCLE NAELKEQMGEAMSDGWEIEEDKEKGEVMVETVVTKEGLSESSLQAEFRKLQGKLKNAH NIINLLKEQLVLSSKEGNSKLTPELLVHLTSTIERINTELVGSPGKHQHQEEGNVTVR PFPRPQSLDLGATFTVDAHQLDNQSQPRDPGPQSAFSLPGSTQHLRSQLSQCKQRYQD LQEKLLLSEATVFAQANELEKYRVMLTGESLVKQDSKQIQVDLQDLGYETCGRSENEA EREETTSPECEEHNSLKEMVLMEGLCSEQGRRGSTLASSSERKPLENQLGKQEEFRVY GKSENILVLRKDIKDLKAQLQNANKVIQNLKSRVRSLSVTSDYSSSLERPRKLRAVGT LEGSSPHSVPDEDEGWLSDGTGAFYSPGLQAKKDLESLIQRVSQLEAQLPKNGLEEKL AEELRSASWPGKYDSLIQDQARELSYLRQKIREGRGICYLITRHAKDTVKSFEDLLRS NDIDYYLGQSFREQLAQGSQLTERLTSKLSTKDHKSEKDQAGLEPLALRLSRELQEKE KVIEVLQAKLDARSLTPSSSHALSDSHRSPSSTSFLSDELEACSDMDIVSEYTHYEEK KASPSHSDSIHHSSHSAVLSSKPSSTSASQGAKAESNSNPISLPTPQNTPKEANQAHS GFHFHSIPKLASLPQAPLPSAPSSFLPFSPTGPLLLGCCETPVVSLAEAQQELQMLQK QLGESASTVPPASTATLLSNDLEADSSYYLNSAQPHSPPRGTIELGRILEPGYLGSSG KWDVMRPQKGSVSGDLSSGSSVYQLNSKPTGADLLEEHLGEIRNLRQRLEESICINDR LREQLEHRLTSTARGRGSTSNFYSQGLESIPQLCNENRVLREDNRRLQAQLSHVSREH SQETESLREALLSSRSHLQELEKELEHQKVERQQLLEDLREKQQEVLHFREERLSLQE NDSRLQHKLVLLQQQCEEKQQLFESLQSELQIYEALYGNSKKGLKGLGLDTSPVMKTP PKLEGDATDGSFANKHGRHVIGHIDDYSALRQQIAEGKLLVKKIVSLVRSACSFPGLE AQGTEVLGSKGIHELRSSTSALHHALEESASLLTMFWRAALPSTHIPVLPGKVGESTE RELLELRTKVSKQERLLQSTTEHLKNANQQKESMEQFIVSQLTRTHDVLKKARTNLEV KSLRALPCTPAL
	SEQ ID NO: 293 7161 bp
NOV103b, CG59773-02 DNA Sequence	GTTGAGGGGGCAATCGGGCACGCTCCTCCCCATGGGTTGCCCATCATGTCTAATGGAT ATCGCACTCTGTCCCAGCACCTCAATGACCTGAAGAAGGAGAACTTCAGCCTCAAGCT GCGCATCTACTTCCTGGAGGAGCGCATGCAACAGAAGTATGAGGCGCAGCCGGGAGGAC ATCTACAAGCGGAACATTGAGCTGAAGGTTGAAGTGGAGAGCTTGAAACGAGAACTCC AGGACAAGAAACAGCATCTGGATAAAACATGGGCTGATGTGGAGAATCTCAACAGTCA GAATGAAGCTGAGCTCCGACGCCAGTTTGAGGAGCGACAGCAGGAGACGGAGCATGTT TATGAGCTCTTGGAGAATAAGATCCAGCTTCTGCAGGAGGAATCCAGGCTAGCAAAGA ATGAAGCTGCGCGGATGGCAGCTCTGGTGGAAGCAGAGAAGGAGTGTAACCTGGAGCT CTCAGAGAAACTGAAGGGAGTCACCAAAAACTGGGAAGATGTACCAGGAGACCAGGTC AAGCCCGACCAATACACTGAGGCCCTGGCCCAGAGGGACAGGAGAATTGAAGAACTGA ATCAGAGCCTGGCTGCCCAGGAGAGGCTTGTAGAACAGCTATCTCGGGAGAAACAACA ACTGCTACATCTGTTGGAGGAGCCAACTAGCATGGAAGTGCAGCCCATGACTGAAGAG TTGCTGAAACAACAAAAGCTGAATTCACATGAGACCACTATAACTCAGCAGTCTGTAT CTGATTCCCACTTGGCAGAACTCCAGGAAAAAATCCAGCAAACAGAGGCCACCAACAA GATTCTTCAAGAGAAACTTAATGAAATGAGCTATGAACTAAAGTGTGCTCAGGAGTCG TCTCAAAAGCAAGATGGTACAATTCAGAACCTCAAGGAAACTCTGAAAAGCAGGGAAC GTGAGACTGAGGAGTTGTACCAGGTAATTGAAGGTCAAAATGACACAATGGCAAAGCT TCGAGAAATGCTGCACCAAAGCCAGCTTGGACAACTTCAGAGCTCAGAGGGTACTTCT CCAGCTCAGCAACAGGTAGCTCTGCTTGATCTTCAGAGTGCTTTATTCTGCAGCCAAC TTGAAATACAGAAGCTCCAGAGGGTGGTACGACAGAAAGAGCGCCAACTGGCTGATGC CAAACAATGTGTGCAATTTGTAGAGGCTGCAGCACACGAGAGTGAACAGCAGAAAGAG GCTTCTTGGAAACATAACCAGGAATTGCGAAAAGCCTTGCAGCAGCTACAAGAAGAAT TGCAGAATAAGAGCCAACAGCTTCGTGCCTGGGAGGCTGAAAAATACAATGAGATTCG AACCCAGGAACAAAACATCCAGCACCTAAACCATAGTCTGAGTCACAAGGAGCAGTTG CTTCAGGAATTTCGGGAGCTCCTACAGTATCGAGATAACTCAGACAAAACCCTTGAAG CAAATGAAATGTTGCTTGAGAAACTTCGCCAGCGAATACATGATAAAGCTGTTGCTCT GGAGCGGGCTATAGATGAAAAATTCTCTGCTCTAGAAGAGAAAGAAAAAGAACTGCGC CAGCTTCGTCTTCGTCTGTGAGAGAGCGAGATCATGACTTAGAGAGACTGCGCGATGTCC TCTCCTCCAATGAAGCTACTATGCAAAGTATGGAGAGTCTCCTGAGGGCCAAAGGCCT GGAAGTGGAACAGTTATCTACTACCTGTCAAAACCTCCAGTGGCTGAAAGAAGAAATG GAAACCAAATTTAGCCGTTGGCAGAAGGAACAAGAGAGTATCATTCAGCAGTTACAGA CGTCTCTTCATGATAGGAACAAAGAAGTGGAGGATCTTGAGTGCAACACTGCTCTGCAA ACTTGGACCAGGGCAGAGTGAGATAGCAGAGGAGCTGTGCCAGCGTCTACAGCGAAAG GAAAGGATGCTGCAGGACCTTCTAAGTGATCGAAATAAACAAGTGCTGGAACATGAAA TGGAGATTCAAGGCCTGCTTCAGTCTGTGAGCACCAGGGAGCAGGAAAGCCAAGCTGC TGCAGAGAAGTTGGTGCAAGCCTTAATGGAAAGAAATTCAGAATTACAGGCCCTGCGC CAATATTTAGGAGGGAGAGACTCCCTGATGTCCCAAGCACCCATCTCTAACCAACAAG CTGAAGTTACCCCCACTGGCCGTCTTGGAAAACAGACTGATCAAGGTTCAATGCAGAT ACCTTCCAGAGATGATAGCACTTCATTGACTGCCAAAGAGGATGTCAGCATACCCAGA TCCACATTAGGAGATTTGGACACAGTTGCAGGGCTGGAAAAAGAACTGAGTAATGCCA

```
AAGAGGAACTTGAACTCATGGCTAAAAAAGAAAGAGAATCACAGATGGAACTTTCTGC
TCTACAGTCCATGATGGCTGTGCAGGAAGAAGAGCTGCAGGTGCAGGCTGCTGATATG
GAGTCTCTGACCAGGAACATACAGATTAAAGAAGATCTCATAAAGGACCTGCAAATGC
AACTGGTTGATCCTGAAGACATACCAGCTATGGAACGCCTGACCCAGGAAGTCTTACT
TCTTCGGGAAAAAGTTGCTTCAGTAGAATCCCAGGGTCAAGAAATTTCAGGAAACCGA
AGACAACAGCAGTTGCTGCTGATGCTAGAAGGACTAGTAGATGAACGGAGTCGGCTCA
ATGAGGCCTTACAAGCAGAGAGAGACAGCTCTATAGCAGTCTGGTGAAGTTCCATGCCCA
TCCAGAGAGCTCTGAGAGAGACCGAACTCTGCAGGTGGAACTGGAAGGGGCTCAGGTG
TTACGCAGTCGGCTAGAAGAAGTTCTTGGAAGAAGCTTGGAGCGCTTAAACAGGCTGG
AGACCCTGGCCGCCATTGGAGGTGCAGCTGCAGGGGATGACACCGAAGATACAAGCAC
TGAGTTCACTGACAGTATTGAGGAGGAGGCTGCACACCATAGTCACCAGCAACTTGTC
AAGGTGGCTTTGGAGAAAAGTCTGGCAACTGTGGAGACCCAGAACCCATCTTTTTCCC
CTCCTTCTCCGATGGGAGGGGACAGTAACAGGTGTCTTCAGGAAGAAATGCTCCACCT
GAGGGCTGAGATCCACCAGCACTTAGAAGAGAAGAGGAAAGCTGAGGAGGAACTGAAG
GAGCTAAAGGCTCAAATTGAGGAAGCAGGATTCTCCTCAGTGTCCCACATCAGGAACA
CCATGCTGAGCCTTTGCCTTGAGAATGCGGAGCTGAAAGAGCAGATGGGAGAAACAAT
GTCTGATGGATGGGAGATCGAGGAAGACAAGGAGAAGGGCGAGGTGATGGTTGAGACT
GTGGTAACCAAAGAGGGTCTGAGTGAGAGTAGCCTTCAGGCTGAGTTCAGAAAGCTCC
AGGGAAAACTGAAGAATGCCCACAATATCATCAACCTCCTCAAAGAACAACTTGTGCT
GAGTAGCAAGGAAGGGAATAGTAAACTTACTCCAGAGCTCCTTGTGCATCTGACCAGC
ACCATCGAAAGAATAAACACAGAACTGGTTGGTTCCCCTGGGAAGCACCAACACCAAG
AGGAGGGGAATGTGACTGTGAGGCCTTTCCCCAGACCCCAGAGCCTTGACCTTGGGGC
TACCTTCACAGTGGATGCCCACCAACAGTTGGATAACCAGTCCCAGCCTCGTGACCCT
GGGCCTCAGCCAGCGTTTAGCCTACCAGGGTCCACCCAGCACCTGCGCTCCCAGCTGT
CACAATGCAAACAACGCTATCAAGATCTCCAGGAGAAGCTGCTGCTATCAGAAGCCAC
TGTCTTTGCTCAGGCTAACGAGCTGGAGAAATACAGAGTTATGCTTAGTGAATCCTTG
GTGAAGCAGGACAGCAAGCAGATCCAGGTGGACTTCCAGGACCTGGGCTATGAGACTT
GTGGCCGAAGCGAGAATGAGGCTGAACGGGAGGAAACCACCAGTCCTGAGTGTGAGGA
GCACAACAGCCTCAAGGAAATGGTCCTGATGGAGGGGCTGTGCTCTGAGCAGGGACGC
CGGGGCTCAACACTGGCTAGTTCCTCTGAGAGGAAGCCCTTGGAGAACCAGCTAGGGA
AGCAGGAAGAGTTCCGGGTATATGGAAAGTCAGAAAACATCTTGGTCCTACGAAAGGA
CATCGAAGATCTGAAGGCCCAGCTGCAGAATGCCAACAAGGTCATTCAAAACCTCAAG
AGCCGGGTCCGGTCCCTCTCAGTTACAAGTGATTATTCGTCTAGTCTGGAAAGACCCC
GGAAGCTGAGAGCTGTTGGCACCTTGGAGGGGTCTTCACCTCATAGTGTCCCTGATGA
GGATGAGGGGTGGCTGTCTGATGGCACTGGGGCTTTCTACTCTCCAGGGCTTCAGGCC
AAAAAGGACCTGGAGAGTCTCATCCAGAGAGTATCCCAGCTGGAGGCCCAGCTCCCAG
AAAATGGACTAGAAGAGAAGCTGGCTGAGGAGCTGAGATCAGCCTCGTGGCCTGGGAA
ATATGATTCCCTGATTCAGGATCAGGCCCGGGAACTGTCTTACCTACGGCAAAAAATA
CGAGAAGGGAGAGGTATTTGTTATCTTATCACCCAGCATGCAAAAGATACAGTAAAAT
CTTTTGAGGATCTCCTAAGGAGCAATGACATTGACTACTACCTGGGACAGAGCTTCCG
GGAGCAACTCGCCCAGGGAAGCCAGCTGACAGAGAGGCTCACCAGCAAACTCAGCACA
GAGGATCATAAAAGTGAGAAAGATCAAGCTGGACTTGAGCCACTGGCCCTCAGGCTCA
GCAGGGAGCTGCAGGAGAAGGAGAAAGTGATTGAAGTCCTGCAGGCCAAGCTGGATGC
TCGGTCCCTCACACCCTCCAGCAGCCGTGCCTTGTCTGACTCCCACCGCTCTCCCAGC
AGCACCTCTTTCCTGTCTGATGAGCTGGAAGCCTGCTCTGACATGGACATAGTCAGCG
AGTACACACACTATGAAGAGAAGAAAGCTTCTCCCAGTCACTCAGGTAGCAGTGCATC
TCAGGGGGCTAAGGCCGAATCCAACAGCAACCCCATCAGCTTGCCAACTCCCCAGAAT
ACCCCCAAGGAGGCCAACCAAGCCCATTCAGGCTTTCATTTTCACTCCATACCCAAGC
TGGCTAGCCTTCCTCAGGCACCATTGCCCTCAGCTCCATCCAGCTTCCTGCCTTTCAG
CCCCACTGGCCCTCCCCTCCTTGGCTGCTGTGAGACACCAGAGGTCCTTGGCTGAG
TCTCAGCAGGAGCTACAGATGCTGCAGAAGCAGTTGGGAGAAAGTAGCACTGTTCCTC
CTGCTTCCACAGCTACATTGCTGAGCAACGACTTGGAAGCCGACTCTTCCTACTACCT
CAACTCTGCCCAGCCTCACTCTCCTCCAAGGGGCACCATAGAACTGGGAAGAATCCTA
GAGCCTGGGTACCTGGGCAGCAGTGGCAAGTGGGATGTGATGAGGCCTCAGAAAGGGA
GTGTATCTGGGGACCTATCCTCAGGCTCCTCTGTGTACCAGCTTAACTCCAAACCCAC
AGGGGCTGACCTGCTGGAAGAGCATCTTGGTGAAATCTGGAACCTGCGCCAGCGCCTG
GAGGAGTCCATCTGCATCAATGACTGCCTACGGGAGCAACTGGAACACCGGCTGACCT
CTACTGCTCGTGGAAGGGGATCCACTTCTAACTTCTACAGTCAGGGCCTGGAGTCCAT
ACCTCAGCTCTGCAATGAGAACAGAGTCCTCAGGGAAGAAAATCGAAGACTTCAGGCT
CAACTGAGTCATGTTTCCAGAGGTCACTCCCAGGAAACAGAAAGCCTGAGGGAGGCTC
TGCTGTCCTCTCGATCCCACCTTCAAGAGCTGGAAAAGGAGCTGGAGCACCAGAAGGT
GGAAAGGCAGCAGCTTTTGGAAGACTTGAGGGAGAAGCAGCAAGAGGTCTTGCATTTC
AGGGAGGAACGTCTTTCCCTCCAGGAAAACGACTCCAGACTGCAGCACAAGCTGGTTC
TCCTGCAGCAACAGTGTGAAGAGAAACAGCAGCTCTTTGAGTCCCTCCAGTCAGAGCT
ACAAATCTACGAGGCACTTTATGGCAATTCCAAGAAGGGGCTGAAAGCTTACAGCCTG
GATGCCTGTCACCAAATCCCTTTGAGCAGTGACCTGAGCCCACCTGGTGGCAGAGGTAC
AAGCTCTGAGAGGGCAGCTGGAGCAGAGCATTCAGGGGAACAATTGTCTGCGACTGCA
GCTGCAACAGCAGCTGGAGAGCGGTGCTGGCAAAGCCAGCCTCAGCCCCTCCTCCATT
AACCAGAACTTCCCAGCCAGCACTGACCCTGGAAACAAGCAGCTGCTCCTCCAAGGTT
CAGCTGTGTCCCCTCCAGTCCGGGATGTTGGTATGAATTCCCCAGCTCTGGTCTTCCC
CAGCTCTGCTTCCTCTACTCCTGGCTCAGATTCAGTTGTGTTGTCATTTTCTTTTTCA
GGCTTGGGTTTGGATACTTCTCCAGTAATGAAGACCCCTCCCAAGCTAGAGGGTGATG
CTACTGATGGCTCCTTTGCCAATAAGCATGGCCGCCATGTCATTGGCCACATTGATGA
CTACAGTGCCCTAAGACAGCAGATTGCGGAGGGCAAGCTGCTGGTCAAAAAGATAGTG
TCTCTTGTGAGATCAGCGTGCAGCTTCCCTGGCCTTGAAGCCCAAGGCACAGAGGGCA
GCAAAGGCATTCATGAGCTTCGGAGCAGCACCAGTGCCCTGCACCATGCCCTAGAGGA
GTCGGCTTCCCTCCTCACCATGTTCTGGAGAGCGGCCCTGCCAAGCACCCACATCCCT
```

405

	GTGCTGCCTGGCAAACAGGGAGAATCAACAGAAAGGGAACTTCTGGAACTGAGAACCA AAGTATCCAAACAGGAGCAGCTCCTTCAGAGCACAACTGAGCATCTGAAGAACGCCAA CCAGCAGAAGGAGAGCATGGAACAGTTCATTGTCAGCGTAACCAGAACACATGATGTT TTAAAGAAGGCAAGGACTAACTTAGAGGTGAAATCCCTAAGGGCTCTGCCGTGTACTC CAGCCTTG**TGA**CCCTTGCCTTCCAGGAACCATGCAAGAAGCGCAGCCACCAGAAGTCC TTAAAACAGCAGGAAAGGTGAGCCTGTCCCCCTTTTGTGCAGCTACCTATCTGCTGAG GAGCATCTGGGCCTCATTCCTCCAAGT
	ORF Start: ATG at 46 | **ORF Stop: TGA at 7027**
	SEQ ID NO: 294 | 2327 aa | MW at 263034.6kD
NOV103b, CG59773-02 Protein Sequence	MSNGYRTLSQHLNDLKKENFSLKLRIYFLEERMQQKYEASREDIYKRNIELKVEVESL KRELQDKKQHLDKTWADVENLNSQNEAELRRQFEERQQETEHVYELLENKIQLLQEES RLAKNEAARMAALVEAEKECNLELSEKLKGVTKNWEDVPGDQVKPDQYTEALAQRDRR IEELNQSLAAQERLVEQLSREKQQLLHLLEEPTSMEVQPMTEELLKQQKLNSHETTIT QQSVSDSHLAELQEKIQQTEATNKILQEKLNEMSYELKCAQESSQKQDGTIQNLKETL KSRERETEELYQVIEGQNDTMAKLREMLHQSQLGQLQSSEGTSPAQQQVALLDLQSAL FCSQLEIQKLQRVVRQKERQLADAKQCVQFVEAAAHESEQQKEASWKHNQELRKALQQ LQEELQNKSQQLRAWEAEKYNEIRTQEQNIQHLNHSLSHKEQLLQEFRELLQYRDNSD KTLEANEMLLEKLRQRIHDKAVALERAIDEKFSALEEKEKELRQLRLAVRERDHDLER LRDVLSSNEATMQSMESLLRAKGLEVEQLSTTCQNLQWLKEEMETKFSRWQKEQESII QQLQTSLHDRNKEVEDLSATLLCKLGPGQSEIAEELCQRLQRKERMLQDLLSDRNKQV LEHEMEIQGLLQSVSTREQESQAAAEKLVQALMERNSELQALRQYLGGRDSLMSQAPI SNQQAEVTPTGRLGKQTDQGSMQIPSRDDSTSLTAKEDVSIPRSTLGDLDTVAGLEKE LSNAKEELELMAKKERESQMELSALQSMMAVQEEELQVQAADMESLTRNIQIKEDLIK DLQMQLVDPEDIPAMERLTQEVLLLREKVASVESQGQEISGNRRQQQLLLMLEGLVDE RSRLNEALQAERQLYSSLVKFHAHPESSERDRTLQVELEGAQVLRSRLEEVLGRSLER LNRLETLAAIGGAAAGDDTEDTSTEFTDSIEEEAAHHSHQQLVKVALEKSLATVETQN PSFSPPSPMGGDSNRCLQEEMLHLRAEIHQHLEEKRKAEEELKELKAQIEEAGFSSVS HIRNTMLSLCLENAELKEQMGETMSDGWEIEEDKEKGEVMVETVVTKEGLSESSLQAE FRKLQGKLKNAHNIINLLKEQLVLSSKEGNSKLTPELLVHLTSTIERINTELVGSPGK HQHQEEGNVTVRPFPRPQSLDLGATFTVDAHQQLDNQSQPRDPGPQPAFSLPGSTQHL RSQLSQCKQRYQDLQEKLLLSEATVFAQANELEKYRVMLSESLVKQDSKQIQVDFQDL GYETCGRSENEAEREETTSPECEEHNSLKEMVLMEGLCSEQGRRGSTLASSSERKPLE NQLGKQEEFRVYGKSENILVLRKDIEDLKAQLQNANKVIQNLKSRVRSLSVTSDYSSS LERPRKLRAVGTLEGSSPHSVPDEDEGWLSDGTGAFYSPGLQAKKDLESLIQRVSQLE AQLPENGLEEKLAEELRSASWPGKYDSLIQDQARELSYLRQKIREGRGICYLITQHAK DTVKSFEDLLRSNDIDYYLGQSFREQLAGSQLTERLTSKLSTEDHKSEKDQAGLEPL ALRLSRELQEKEKVIEVLQAKLDARSLTPSSSRALSDSHRSPSSTSFLSDELEACSDM DIVSEYTHYEEKKASPSHSGSSASQGAKAESNSNPISLPTPQNTPKEANQAHSGFHFH SIPKLASLPQAPLPSAPSSFLPFSPTGPPLLGCCETPEVSLAESQQELQMLQKQLGES STVPPASTATLLSNDLEADSSYYLNSAQPHSPPRGTIELGRILEPGYLGSSGKWDVMR PQKGSVSGDLSSGSSVYQLNSKPTGADLLEEHLGEIWNLRQRLEESICINDCLREQLE HRLTSTARGRGSTSNFYSQGLESIPQLCNENRVLREENRRLQAQLSHVSRGHSQETES LREALLSSRSHLQELEKELEHQKVERQQLLEDLREKQQEVLHFREERLSLQENDSRLQ HKLVLLQQQCEEKQQLFESLQSELQIYEALYGNSKKGLKAYSLDACHQIPLSSDLSHL VAEVQALRGQLEQSIQGNNCLRLQLQQQLESGAGKASLSPSSINQNFPASTDPGNKQL LLQGSAVSPPVRDVGMNSPALVFPSSASSTPGSDSVVLSFSFSGLGLDTSPVMKTPPK LEGDATDGSFANKHGRHVIGHIDDYSALRQQIAEGKLLVKKIVSLVRSACSFPGLEAQ GTEGSKGIHELRSSTSALHHALEESASLLTMFWRAALPSTHIPVLPGKQGESTERELL ELRTKVSKQEQLLQSTTEHLKNANQQKESMEQFIVSVTRTHDVLKKARTNLEVKSLRA LPCTPAL
	SEQ ID NO: 295 | 7084 bp
NOV103c, CG59773-03 DNA Sequence	GTTGAGGGGGCAATCGGGCACGCTCCTCCCCATGGGTTGCCCATCATGTCTAATGGAT ATCGCACTCTGTCCCAGCACCTCAATGACCTGAAGAAGGAGAACTTCAGCCTCAAGCT GCTCATCTACTTCCTGGAGGAGCGCATGCAACAGAAGT**ATG**AGGCCAGCCGGGAGGAC ATCTACAAGCGGGGGTGATGTGGAGAATCTCAACAGTCAGAATGAAGCTGAGCTCCGA CGCCAGTTTGAGGAGCGACAGCAGGAGACGGAGCATGTTTATGAGCTCTTGGAGAATA AGATCCAGCTTCTGCAGGAGGAATCCAGGCTAGCAAAGAATGAAGCTGCGCGGATGGC AGCTCTGGTGGAAGCAGAGAAGGAGTGTAACCTGGAGCTCTCAGAGAAACTGAAGGGA GTCACCAAAAACTGGGAAGATGTACCAGGAGACCAGGTCAAGCCCGACCAATACACTG AGACCCTGGCCCAGAGGGACAAGAGAATTGAAGAACTGAATCAGAGCCTGGCTGCCCA GGAGAGGCTTGTAGAACAGCTATCTCGGGAGAAACAACAACTGCTACATCTGTTGGAG GAGCCAACTAGCATGGAAGTGCAGCCCATGACTGAAGAGTTGCTGAAACAACAAAAGC TGAATTCACATGAGACCACTATAACTCAGCAGTCTGTATCTGATTCCCACTTGGCAGA ACTCCAGGAAAAAATCCAGCAAACAGAGGCCACCAACAAGATTCTTCAAGAGAAACTT AATGAAATGAGCTATGAACTAAAGTGTGCTCAGGAGTCGTCTCAAAAGCAAGATGGTA CAATTCAGAACCTCAAGGAAACTCTGAAAAGCAGGGAACGTGAGACTGAGGAGTTGTA CCAGGTAATTGAAGGTCAAAATGACACAATGGCAAAGCTTCGAGAAATGCTGCACCAA AGCCAGCTTGGACAACTTCAGAGCTCAGAGGGTACTTCTCCAGCTCAGCAACAGGTAG CTCTGCTTGATCTTCAGAGTGCTTTATTCTGCAGCCAACTTGAAATACAGAAGCTCCA GAGGGTGGTACGACAGAAAGAGCGCCAACTGGCTGATGCCAAACAATGTGTGCAATTT GTAGAGGCTGCAGCACACGAGAGTGAACAGCAGAAAGAGGCTTCTTGGAAACATAACC AGGAATTGCGAAAAGCCTTGCAGCAGCTACAAGAAGAATTGCAGAATAAGAGCCAACA GCTTCGTGCCTGGGAGGCTGAAAAATACAATGAGATTCGAACCCAGGAACAAAACATC

```
CAGCACCTAAACCATAGTCTGAGTCACAAGGAGCAGTTGCTTCAGGAATTTCGGGAGC
TCCTACAGTATCGAGATAACTCAGACAAAACCCTTGAAGCAAATGAAATGTTGCTTGA
GAAACTTCGCCAGCGAATACATGATAAAGCTGTTGCTCTGGAGCGGGCTATAGATGAA
AAATTCTCTGCTCTAGAAGAGAAAGAAAAAGAACTGCGCCAGCTTCGTCTTGCTGTGA
GAGAGCGAGATCATGACTTAGAGAGACTGCGCGATGTCCTCTCCTCCAATGAAGCTAC
TATGCAAAGTATGGAGAGTCTCCTGAGGGCCAAAGGCCTGGAAGTGGAACAGTTATCT
ACTACCTGTCAAAACCTCCAGTGGCTGAAAGAAGAAATGGAAACCAAATTTAGCCGTT
GGCAGAAGGAACAAGAGAGTATCATTCAGCAGTTACAGACGTCTCTTCATGATAGGAA
CAAAGAAGTGGAGGATCTTAGTGCAACACTGCTCTGCAAACTTGGACCAGGGCAGAGT
GAGATAGCAGAGGAGCTGTGCCAGCGTCTACAGCGAAAGGAAAGGATGCTGCAGGACC
TTCTAAGTGATCGAAATAAACAAGTGCTGGAACATGAAATGGAGATTCAAGGCCTGCT
TCAGTCTGTGAGCACCAGGGAGCAGGAAAGCCAAGCTGCTGCAGAGAAGTTGGTGCAA
GCCTTAATGGAAAGAAATTCAGAATTACAGGCCCTGCGCCAATATTTAGGAGGGAGAG
ACTCCCTGATGTCCCAAGCACCCATCTCTAACCAACAAGCTGAAGTTACCCCCACTGG
CCGTCTTGGAAAACAGACTGATCAAGGTTCAATGCAGATACCTTCCAGAGATGATAGC
ACTTCATTGACTGCCAAAGAGGATGTCAGCATACCCAGATCCACATTAGGAGATTTGG
ACACAGTTGCAGGGCTGGAAAAAGAACTGAGTAATGCCAAAGAGGAACTTGAACTCAT
GGCTAAAAAAGAAAGAGAATCACAGATGGAACTTTCTGCTCTACAGTCCATGATGGCT
GTGCAGGAAGAAGAGCTGCAGGTGCAGGCTGCTGATATGGAGTCTCTGACCAGGAACA
TACAGATTAAAGAAGATCTCATAAAGGACCTGCAAATGCAACTGGTTGATCCTGAAGA
CATACCAGCTATGGAACGCCTGACCCAGGAAGTCTTACTTCTTCGGGAAAAAGTTGCT
TCAGTAGAATCCCAGGGTCAAGAAATTTCAGGAAACCGAAGACAACAGCAGTTGCTGC
TGATGCTAGAAGGACTAGTAGATGAACGGAGTCGGCTCAATGAGGCCTTACAAGCAGA
GAGACAGCTCTATAGCAGTCTGGTGAAGTTCCATGCCCATCCAGAGAGCTCTGAGAGA
GACCGAACTCTGCAGGTGGAACTGGAAGGGGCTCAGGTGTTACGCAGTCGGCTAGAAG
AAGTTCTTGGAAGAAGCTTGGAGCGCTTAAACAGGCTGGAGACCCTGGCCGCCATTGG
AGGTGCAGCTGCAGGGGATGACACCGAAGATACAAGCACTGAGTTCACTGACAGTATT
GAGGAGGAGGCTGCACACCATAGTCACCAGCAACTTGTCAAGGTGGCTTTGGAGAAAA
GTCTGGCAACTGTGGAGACCCAGAACCCATCTTTTTCCCCTCCTTCTCCGATGGGGAGG
GGACAGTAACAGGTGTCTTCAGGAAGAAATGCTCCACCTGAGGGCTGAGATCCACCAG
CACTTAGAAGAGAAGAGGAAAGCTGAGGAGGAACTGAAGGAGCTAAAGGCTCAAATTG
AGGAAGCAGGATTCTCCTCAGTGTCCCACATCAGGAACACCATGCTGAGCCTTTGCCT
TGAGAATGCGGGAGCTGAAAGAGCAGATGGGAGAAACAATGTCTGATGGATGGGAGATC
GAGGAAGACAAGGAGAAGGGCGAGGTGATGGTTGAGACTGTGGTAACCAAAGAGGGTC
TGAGTGAGAGTAGCCTTCAGGCTGAGTTCAGAAAGCTCCAGGGAAAACTGAAGAATGC
CCACAATATCATCAACCTCCTCAAAGAACAACTTGTGCTGAGTAGCAAGGAAGGGAAT
AGTAAACTTACTCCAGAGCTCCTTGTGCATCTGACCAGCACCATCGAAAGAATAAACA
CAGAACTGGTTGGTTCCCCTGGGAAGCACCAACACCAAGAGGAGGGGAATGTGACTGT
GAGGCCTTTCCCCAGACCCCAGAGCCTTGACCTTGGGGCTACCTTCACAGTGGATGCC
CACCAACAGTTGGATAACCAGTCCCAGCCTCGTGACCCTGGGCCTCAGCCAGCGTTTA
GCCTACCAGGGTCCACCCAGCACCTGCGCTCCCAGCTGTCACAATGCAAACAACGCTA
TCAAGATCTCCAGGAGAAGCTGCTGCTATCAGAAGCCACTGTCTTTGCTCAGGCTAAC
GAGCTGGAGAAATACAGAGTTATGCTTAGTGAATCCTTGGTGAAGCAGGACAGCAAGC
AGATCCAGGTGGACTTCCAGGACCTGGGCTATGAGACTTGTGGCCGAAGCGAGAATGA
GGCTGAACGGGAGGAAACCACCAGTCCTGAGTGTGAGGAGCACAACAGCCTCAAGGAA
ATGGTCCTGATGGAGGGGCTGTGCTCTGAGCAGGGACGCCGGGGCTCAACACTGGCTA
GTTCCTCTGAGAGGAAGCCCTTGGAGAACCAGCTAGGGAAGCAGGAAGAGTTCCGGGT
ATATGGAAAGTCAGAAAACATCTTGGTCCTACGAAAGGACATCGAAGATCTGAAGGCC
CAGCTGCAGAATGCCAACAAGGTCATTCAAAACCTCAAGAGCCGGGTCCGGTCCCTCT
CAGTTACAAGTGATTATTCGTCTAGTCTGGAAAGACCCCGGAAGCTGAGAGCTGTTGG
CACCTTGGAGGGGTCTTCACCTCATAGTGTCCCTGATGAGGATGAGGGGTGGCTGTCT
GATGGCACTGGGGCTTTCTACTCTCCAGGGCTTCAGGCCAAAAAGGACCTGGAGAGTC
TCATCCAGAGAGTATCCCAGCTGGAGGCCCAGCTCCCAGAAAATGGACTAGAAGAGAA
GCTGGCTGAGGAGCTGAGATCAGCCTCGTGGCCTGGGAAATATGATTCCCTGATTCAG
GATCAGGCCCGGGAACTGTCTTACCTACGGCAAAAAATACGAGAAGGGAGAGGTATTT
GTTATCTTATCACCCAGCATGCAAAAGATACAGTAAAATCTTTTGAGGATCTCCTAAG
GAGCAATGACATTGACTACTACCTGGGACAGAGCTTCCGGGAGCAACTCGCCCAGGGA
AGCCAGCTGACAGAGAGGCTCACCAGCAAACTCAGCACAGAGGATCATAAAAGTGAGA
AAGATCAAGCTGGACTTGAGCCACTGGCCCTCAGGCTCAGCAGGGAGCTGCAGGAGAA
GGAGAAAGTGATTGAAGTCCTGCAGGCCAAGCTGGATGCTCGGTCCCTCACACCCTCC
AGCAGCCGTGCCTTGTCTGACTCCCACCGCTCTCCCAGCAGCACCTCTTTCCTGTCTG
ATGAGCTGGAAGCCTGCTCTGACATGGACATAGTCAGCGAGTACACACACTATGAAGA
GAAGAAAGCTTCTCCCAGTCACTCAGGTAGCAGTGCATCTCAGGGGGCTAAGGCCGAA
TCCAACAGCAACCCCATCAGCTTGCCAACTCCCCAGAATACCCCCAAGGAGGCCAACC
AAGCCCATTCAGGCTTTCATTTTCACTCCATACCCAAGCTGGCTAGCCTTCCTCAGGC
ACCATTGCCCTCAGCTCCATCCAGCTTCCTGCCTTTCAGCCCCACTGGCCCTCCCCTC
CTTGGCTGCTGTGAGACACCAGAGGTCTCCTTGGCTGAGTCTCAGCAGGAGCTACAGA
TGCTGCAGAAGCAGTTGGGAGAAAGTAGCACTGTTCCTCCTGCTTCCACAGCTACATT
GCTGAGCAACGACTTGGAAGCCGACTCTTCCTACTACCTCAACTCTGCCCAGCCTCAC
TCTCCTCCAAGGGGCACCATAGAACTGGGAAGAATCCTAGAGCCTGGGTACCTGGGCA
GCAGTGGCAAGTGGGATGTGATGAGGCCTCAGAAAGGGAGTGTATCTGGGGACCTATC
CTCAGGCTCCTCTGTGTACCAGCTTAACTCCAAACCCACAGGGGCTGACCTGCTGGAA
GAGCATCTTGGTGAAATCTGGAACCTGCGCCAGCGCCTGGAGGAGTCCATCTGCATCA
ATGACTGCCTACGGGAGCAACTGGAACACCGGCTGACCTCTACTGCTCGTGGAAGGGG
ATCCACTTCTAACTTCTACAGTCAGGGCCTGGAGTCCATACCTCAGCTCTGCAATGAG
AACAGAGTCCTCAGGGAAGAAAATCGAAGACTTCAGGCTCAACTGAGTCATGTTTCCA
GAGGTCACTCCCAGGAAACAGAAAGCCTGAGGGAGGCTCTGCTGTCCTCTCGATCCCA
```

407

	CCTTCAAGAGCTGGAAAAGGAGCTGGAGCACCAGAAGGTGGAAAGGCAGCAGCTTTTG GAAGACTTGAGGGAGAAGCAGCAAGAGGTCTTGCATTTCAGGGAGGAACGTCTTTCCC TCCAGGAAAACGACTCCAGACTGCAGCACAAGCTGGTTCTCCTGCAGCAACAGTGTGA AGAGAAACAGCAGCTCTTTGAGTCCCTCCAGTCAGAGCTACAAATCTACGAGGCACTT TATGGCAATTCCAAGAAGGGGCTGAAAGCTTACAGCCTGGATGCCTGTCACCAAATCC CTTTGAGCAGTGACCTGAGCCACCTGGTGGCAGAGGTACAAGCTCTGAGAGGGCAGCT GGAGCAGAGCATTCAGGGGAACAATTGTCTGCGACTGCAGCTGCAACAGCAGCTGGAG AGCGGTGCTGGCAAAGCCAGCCTCAGCCCCTCCTCCATTAACCAGAACTTCCCAGCCA GCACTGACCCTGGAAACAAGCAGCTGCTCCTCCAAGGTTCAGCTGTGTCCCCTCCAGT CCGGGATGTTGGTATGAATTCCCCAGCTCTGGTCTTCCCCAGCTCTGCTTCCTCTACT CCTGGCTCAGATTCAGTTGTGTTGTCATTTTCTTTTTCAGGCTTGGGTTTGGATACTT CTCCAGTAATGAAGACCCCTCCCAAGCTAGAGGGTGATGCTACTGATGGCTCCTTTGC CAATAAGCATGGCCGCCATGTCATTGGCCACATTGATGACTACAGTGCCCTAAGACAG CAGATTGCGGAGGGCAAGCTGCTGGTCAAAAAGATAGTGTCTCTTGTGAGATCAGCGT GCAGCTTCCCTGGCCTTGAAGCCCAAGGCACAGAGGGCAGCAAAGGCATTCATGAGCT TCGGAGCAGCACCAGTGCCCTGCACCATGCCCTAGAGGAGTCGGCTTCCCTCCTCACC ATGTTCTGGAGAGCGGCCCTGCCAAGCACCCACATCCCTGTGCTGCCTGGCAAACAGG GAGAATCAACAGAAGGGAACTTCTGGAACTGAGAACCAAAGTATCCAAACAGGAGCA GCTCCTTCAGAGCACAACTGAGCATCTGAAGAACGCCAACCAGCAGAAGGAGAGCATG GAACAGTTCATTGTCAGCGTAACCAGAACACATGATGTTTTAAAGAAGGCAAGGACTA ACTTAGAGGTGAAATCCCTAAGGGCTCTGCCCGTGTACTCCAGCCTTGTGACCCTTGCC TTCCAGGAACCATGCAAGAAGCGCAGCCACCAGAAGTCCTTAAAACAGCAGGAAAGGT GAGCCTGTCCCCCTTTTGTGCAGCTACCTATCTGCTGAGGAGCATCTGGGCCTCATTC CTCCAAGT
	ORF Start: ATG at 155	**ORF Stop: TGA at 6950**	
	SEQ ID NO: 296	**2265 aa**	**MW at 255081.5kD**
NOV103c, CG59773-03 Protein Sequence	MRPAGRTSTSGGDVENLNSQNEAELRRQFEERQQETEHVYELLENKIQLLQEESRLAK NEAARMAALVEAEKECNLELSEKLKGVTKNWEDVPGDQVKPDQYTETLAQRDKRIEEL NQSLAAQERLVEQLSREKQQLLHLLEEPTSMEVQPMTEELLKQQKLNSHETTITQQSV SDSHLAELQEKIQQTEATNKILQEKLNEMSYELKCAQESSQKQDGTIQNLKETLKSRE RETEELYQVIEGQNDTMAKLREMLHQSQLGQLQSSEGTSPAQQQVALLDLQSALFCSQ LEIQKLQRVVRQKERQLADAKQCVQFVEAAAHESEQQKEASWKHNQELRKALQQLQEE LQNKSQQLRAWEAEKYNEIRTQEQNIQHLNHSLSHKEQLLQEFRELLQYRDNSDKTLE ANEMLLEKLRQRIHDKAVALERAIDEKFSALEEKEKELRQLRLAVRERDHDLERLRDV LSSNEATMQSMESLLRAKGLEVEQLSTTCQNLQWLKEEMETKFSRWQKEQESIIQQLQ TSLHDRNKEVEDLSATLLCKLGPGQSEIAEELCQRLQRKERMLQDLLSDRNKQVLEHE MEIQGLLQSVSTREQESQAAAEKLVQALMERNSELQALRQYLGGRDSLMSQAPISNQQ AEVTPTGRLGKQTDQGSMQIPSRDDSTSLTAKEDVSIPRSTLGDLDTVAGLEKELSNA KEELELMAKKERESQMELSALQSMMAVQEEELQVQAADMESLTRNIQIKEDLIKDLQM QLVDPEDIPAMERLTQEVLLLREKVASVESQGQEISGNRRQQQLLLMLEGLVDERSRL NEALQAERQLYSSLVKFHAHPESSERDRTLQVELEGAQVLRSRLEEVLGRSLERLNRL ETLAAIGGAAAGDDTEDTSTEFTDSIEEEAAHHSHQQLVKVALEKSLATVETQNPSFS PPSPMGGDSNRCLQEEMLHLRAEIHQHLEEKRKAEEELKELKAQIEEAGFSSVSHIRN TMLSLCLENAELKEQMGETMSDGWEIEEDKEKGEVMVETVVTKEGLSESSLQAEFRKL QGKLKNAHNIINLLKEQLVLSSKEGNSKLTPELLVHLTSTIERINTELVGSPGKHQHQ EEGNVTVRPFPRPQSLDLGATFTVDAHQQLDNQSQPRDPGPQPAFSLPGSTQHLRSQL SQCKQRYQDLQEKLLLSEATVFAQANELEKYRVMLSESLVKQDSKQIQVDFQDLGYET CGRSENEAEREETTSPECEEHNSLKEMVLMEGLCSEQGRRGSTLASSSERKPLENQLG KQEEFRVYGKSENILVLRKDIEDLKAQLQNANKVIQNLKSRVRSLSVTSDYSSSLERP RKLRAVGTLEGSSPHSVPDEDEGWLSDGTGAFYSPGLQAKKDLESLIQRVSQLEAQLP ENGLEEKLAEELRSASWPGKYDSLIQDQARELSYLRQKIREGRGICYLITQHAKDTVK SFEDLLRSNDIDYYLGQSFREQLAQGSQLTERLTSKLSTEDHKSEKDQAGLEPLALRL SRELQEKEKVIEVLQAKLDARSLTPSSSRALSDSHRSPSSTSFLSDELEACSDMDIVS EYTHYEEKKASPSHSGSSASQGAKAESNSNPISLPTPQNTPKEANQAHSGFHFHSIPK LASLPQAPLPSAPSSFLPFSPTGPPLLGCCETPEVSLAESQQELQMLQKQLGESSTVP PASTATLLSNDLEADSSYYLNSAQPHSPPRGTIELGRILEPGYLGSSGKWDVMRPQKG SVSGDLSSGSSVYQLNSKPTGADLLEEHLGEIWNLRQRLEESICINDCLREQLEHRLT STARGRGSTSNFYSQGLESIPQLCNENRVLREENRRLQAQLSHVSRGHSQETESLREA LLSSRSHLQELEKELEHQKVERQQLLEDLREKQQEVLHFREERLSLQENDSRLQHKLV LLQQQCEEKQQLFESLQSELQIYEALYGNSKKGLKAYSLDACHQIPLSSDLSHLVAEV QALRGQLEQSIQGNNCLRLQLQQQLESGAGKASLSPSSINQNFPASTDPGNKQLLLQG SAVSPPVRDVGMNSPALVFPSSASSTPGSDSVVLSFSFSGLGLDTSPVMKTPPKLEGD ATDGSFANKHGRHVIGHIDDYSALRQQIAEGKLLVKKIVSLVRSACSFPGLEAQGTEG SKGIHELRSSTSALHHALEESASLLTMFWRAALPSTHIPVLPGKQGESTERELLELRT KVSKQEQLLQSTTEHLKNANQQKESMEQFIVSVTRTHDVLKKARTNLEVKSLRALPCT PAL

Sequence comparison of the above protein sequences yields the following sequence
relationships shown in Table 103B.

Table 103B. Comparison of NOV103a against NOV103b through NOV103c.		
Protein Sequence	NOV103a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV103b	365..2196 202..2016	1510/1834 (82%) 1518/1834 (82%)
NOV103c	365..2196 140..1954	1510/1834 (82%) 1518/1834 (82%)

Further analysis of the NOV103a protein yielded the following properties shown in Table 103C.

Table 103C. Protein Sequence Properties NOV103a	
PSort analysis:	0.5855 probability located in mitochondrial matrix space; 0.4200 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.2957 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 39 and 40

A search of the NOV103a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 103D.

Table 103D. Geneseq Results for NOV103a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV103a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY71159	Human phosphodiesterase interacting protein, myomegalin - Homo sapiens, 2517 aa. [WO200027861-A1, 18-MAY-2000]	1..2196 1..2204	2193/2204 (99%) 2193/2204 (99%)	0.0
AAM40183	Human polypeptide SEQ ID NO 3328 - Homo sapiens, 1883 aa. [WO200153312-A1, 26-JUL-2001]	635..2196 1..1570	1557/1570 (99%) 1559/1570 (99%)	0.0
AAY71158	Rat phosphodiesterase interacting protein, myomegalin - Rattus sp, 2326 aa. [WO200027861-A1, 18-MAY-2000]	365..2197 202..2017	1433/1837 (78%) 1572/1837 (85%)	0.0

409

AAY67600	Human adipose tissue protein #3 - Homo sapiens, 944 aa. [JP2000037190-A, 08-FEB-2000]	1..934 1..934	925/934 (99%) 927/934 (99%)	0.0
AAU01768	Human secreted protein #47 - Homo sapiens, 934 aa. [WO200123546-A1, 05-APR-2001]	365..1102 197..934	730/738 (98%) 733/738 (98%)	0.0

In a BLAST search of public sequence databases, the NOV103a protein was found to have homology to the proteins shown in the BLASTP data in Table 103E.

Table 103E. Public BLASTP Results for NOV103a				
Protein Accession Number	Protein/Organism/Length	NOV103a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O75042	KIAA0454 PROTEIN - Homo sapiens (Human), 1882 aa (fragment).	636..2196 1..1569	1558/1569 (99%) 1558/1569 (99%)	0.0
Q9WUJ3	MYOMEGALIN - Rattus norvegicus (Rat), 2324 aa.	365..2197 202..2015	1444/1838 (78%) 1581/1838 (85%)	0.0
O75065	KIAA0477 PROTEIN - Homo sapiens (Human), 1132 aa.	1..1132 1..1132	1132/1132 (100%) 1132/1132 (100%)	0.0
Q25893	LIVER STAGE ANTIGEN - Plasmodium falciparum (isolate NF54), 1909 aa.	356..1459 605..1651	243/1129 (21%) 488/1129 (42%)	4e-35
Q13439	Golgi autoantigen, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa golgin) (Golgin-245) (72.1 protein) - Homo sapiens (Human), 2230 aa.	229..1749 267..1814	349/1638 (21%) 679/1638 (41%)	4e-34

PFam analysis predicts that the NOV103a protein contains the domains shown in the Table 103F.

5

410

Table 103F. Domain Analysis of NOV103a			
Pfam Domain	NOV103a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Somatomedin_B: domain 1 of 1	150..189	14/47 (30%) 25/47 (53%)	7.6
recA: domain 1 of 1	621..650	8/30 (27%) 22/30 (73%)	8.1
Ribosomal_L10: domain 1 of 1	604..695	20/109 (18%) 59/109 (54%)	9.9
Dishevelled: domain 1 of 1	844..914	19/74 (26%) 37/74 (50%)	2.7
Transposase_22: domain 1 of 1	1135..1416	71/376 (19%) 127/376 (34%)	4.6
Phe_tRNA-synt_N: domain 1 of 1	2079..2152	13/79 (16%) 49/79 (62%)	4.9

Example 104.

The NOV104 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 104A.

Table 104A. NOV104 Sequence Analysis			
	SEQ ID NO: 297	736 bp	
NOV104a, CG57460-01 DNA Sequence	AAAGCACCCGAGATGACCCCGGCTCCTCCACCAGGAGCGCGGCCGGGCGCGGCGTCCC TAGCGGGCTTCGCCGGGGTGGCGTCTCTGGGGCCTGGGGACCCCCGCCGCGCCGCTGA CCCGCGCCCTCTGCCCCCAGCGCTGTGCTTCGCCGTGAGCCGCTCGCTGCTGCTGACG TGCCTGGTGCCGGCCGCGCTGCTGGGCCTGCGCTACTACTACAGCCGCAAGGTGATCC GCGCCTACCTGGAGTGCGCGCTGCACACGGACATGGCGGACATCGAGCAGTACTACAT GAAGCCGCCCGGTGTGTCCCTGACCGCCCTATCCCCTGCAGGCTCCTGCTTCTGGGTG GCCGTGCTGGATGGCAACGTGGTGGGCATTGTGGCTGCACGGGCCCACGAGGAGGACA ACACGGTGGAGCTGCTGCGGATGTCTGTGGACTCACGTTTCCGAGGCAAGGGCATCGC CAAGGCGCTGGGCCGGAAGGTGCTGGAGTTCGCCGTGGTGCACAACTACTCCGCGGTG GTGCTGGGCACGACGGCCGTCAAGGTGGCCGCCCACAAGCTCTACGAGTCGCTGGGCT TCAGACACATGGGCGCCAGTGACCACTACGTGCTGCCGGGCATGACCCTCTCGCTGGC TGAGCGCCTCTTCTTCCAGGTCCGCTACCACCGCTACCGCCTGCAGCTGCGCGAGGAG TGACCGCCGCCGCTCGCCCGCCCGCCCCCCCCGGCCGCCCT		
	ORF Start: ATG at 13	ORF Stop: TGA at 697	
	SEQ ID NO: 298	228 aa	MW at 24767.5kD
NOV104a, CG57460-01 Protein Sequence	MTPAPPPGARPGAASLAGFAGVASLGPGDPRRAADPRPLPPALCFAVSRSLLLTCLVP AALLGLRYYYSRKVIRAYLECALHTDMADIEQYYMKPPGVSLTALSPAGSCFWVAVLD GNVVGIVAARAHEEDNTVELLRMSVDSRFRGKGIAKALGRKVLEFAVVHNYSAVVLGT TAVKVAAHKLYESLGFRHMGASDHYVLPGMTLSLAERLFFQVRYHRYRLQLREE		

411

Further analysis of the NOV104a protein yielded the following properties shown in Table 104B.

Table 104B. Protein Sequence Properties NOV104a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 64 and 65

A search of the NOV104a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 104C.

Table 104C. Geneseq Results for NOV104a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV104a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB19986	Human camello 3 (Hcml3) protein (partial) - Homo sapiens, 144 aa. [WO200077024-A1, 21-DEC-2000]	42..195 1..144	144/154 (93%) 144/154 (93%)	7e-76
AAB19985	Human camello 2 (Hcml2) protein - Homo sapiens, 227 aa. [WO200077024-A1, 21-DEC-2000]	47..200 56..203	63/158 (39%) 92/158 (57%)	1e-21
AAB19984	Human camello 1 (Hcml1) protein - Homo sapiens, 227 aa. [WO200077024-A1, 21-DEC-2000]	41..196 50..199	60/160 (37%) 88/160 (54%)	7e-20
AAY57959	Human TSC501 protein SEQ ID NO:1 - Homo sapiens, 227 aa. [JP11332579-A, 07-DEC-1999]	41..196 50..199	59/160 (36%) 87/160 (53%)	4e-19
AAB19987	Mouse camello 1 (Mcml1) protein - Mus sp, 222 aa. [WO200077024-A1, 21-DEC-2000]	41..194 50..197	63/158 (39%) 87/158 (54%)	1e-18

In a BLAST search of public sequence databases, the NOV104a protein was found to have homology to the proteins shown in the BLASTP data in Table 104D.

412

Table 104D. Public BLASTP Results for NOV104a

Protein Accession Number	Protein/Organism/Length	NOV104a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UHF3	PUTATIVE N-ACETYLTRANSFERASE CAMELLO 2 - Homo sapiens (Human), 227 aa.	47..200 56..203	63/158 (39%) 92/158 (57%)	5e-21
Q9UHE5	PUTATIVE N-ACETYLTRANSFERASE CML1 - Homo sapiens (Human), 227 aa.	41..196 50..199	60/160 (37%) 88/160 (54%)	3e-19
Q9UQ17	GLA PROTEIN - Homo sapiens (Human), 227 aa.	41..196 50..199	60/160 (37%) 88/160 (54%)	3e-19
Q96QI8	KIDNEY-AND LIVER-SPECIFIC GENE - Homo sapiens (Human), 227 aa.	41..196 50..199	59/160 (36%) 87/160 (53%)	1e-18
O75839	TSC501 PROTEIN - Homo sapiens (Human), 227 aa.	41..196 50..199	59/160 (36%) 87/160 (53%)	1e-18

PFam analysis predicts that the NOV104a protein contains the domains shown in the Table 104E.

Table 104E. Domain Analysis of NOV104a

Pfam Domain	NOV104a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Acetyltransf: domain 1 of 1	111..191	28/82 (34%) 64/82 (78%)	2.2e-17

Example 105.

The NOV105 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 105A.

Table 105A. NOV105 Sequence Analysis

	SEQ ID NO: 299	1230 bp
NOV105a, CG57464-01 DNA Sequence	CTTCCCGGCGGCTGCGCGATGGACAGCCCCGAGGTGACCTTCACTCTCGCCTATCTGG TGTTCGCCGTGTGCTTCGTGTTCACGCCCAACGAGTTCCACGCGGCGGGGCTCACGGT GCAGAACCTGCTGTCGGGCTGGCTGGGCAGCGAGGACGCCGCCTTCGTGCCCTTCCAC TTGCGCCGCACGGCCGCCACGCTGTTGTGCCACTCGCTGCTGCCGCTCGGTGAGGCTG CTCGGGCCGGCCGGCCGCATCCTCTCCTGCGCAGGGCTTGCTGGGAGGTCAGGAGGAG GCCTCCGCCAGCTCCCCGAGGCCCCGAAAGCGCCTGGGCGCAGCTGGGGAGAGGCGCC GGTCCTCATCCAGAGGGACCGCGGCGTGGGCTGAGCGCGCTTAGGGGTGCCGCCGGCC TGGCCTGGCGGCTCTTCCTGCTGCTGGCCGTGACCCTCCCCTCCATCGCCTGCATCCT	

413

	GATCTACTACTGGTCCCGTGACCGGTGGGCCTGCCACCCACTGGCGCGCACCCTGGCC CTCTACGCCCTCCCACAGTCTGGCTGGCAGGCTGTTGCCTCCTCTGTCAACACTGAGT TCCGGCGGATTGACAAGTTTGCCACCGGTGCACCAGGTGCCCGTGTGATTGTGACAGA CACGTGGGTGATGAAGGTAACCACCTACCGAGTGCACGTGGCCCAGCAGCAGGACGTG CACCTGACTGTGACGGAGTCTCGGCAGCATGAGCTCTCGCCAGACTCGAACTTGCCCG TGCAGCTCCTCACCATCCGTGTGGCCAGCACCAACCCTGCTGTGCAGGCCTTTGACAT CAGGCTGAACTCCACTGAGTACGGGGAGCTCTGCGAGAAGCTCCGGGCACCCATCCGC AGGGCAGCCCATGTGGTCATCCACCAGAGCCTGGGCGACCTGTTCCTGGAGACATTTG CCTCCCTGGTAGAGGTCAACCCGGCCTACTCAGTGCCCAGCAGCCAGGTGGGGGGCCT GGAGGCCTGCATAGGCTGCATGCAGACACGTGCCAGCGTGAAGCTGGTGAAGACCTGC CAGGAGGCAGCCACAGGCGAGTGCCAGCAGTGTTACTGCCGCCCCATGTGGTGCCTCA CCTGCATGGGCAAGTGGTTCGCCAGCCGCCAGGACCCCCTGCGCCCTGACACCTGGCT GGCCAGCCGCGTGCCCTGCCCCACCTGCCGCGCACGCTTCTGCATCCTGGATGTGTGC ACCGTGCGCTGA
	ORF Start: ATG at 19 ORF Stop: TGA at 1228
	SEQ ID NO: 300 403 aa MW at 44585.0kD
NOV105a, CG57464-01 Protein Sequence	MDSPEVTFTLAYLVFAVCFVFTPNEFHAAGLTVQNLLSGWLGSEDAAFVPFHLRRTAA TLLCHSLLPLGEAARAGRPHPLLRRACWEVRRRPPPAPRGPESAWAQLGRGAGPHPEG PRRGLSALRGAAGLAWRLFLLLAVTLPSIACILIYYWSRDRWACHPLARTLALYALPQ SGWQAVASSVNTEFRRIDKFATGAPGARVIVTDTWVMKVTTYRVHVAQQQDVHLTVTE SRQHELSPDSNLPVQLLTIRVASTNPAVQAFDIRLNSTEYGELCEKLRAPIRRAAHVV IHQSLGDLFLETFASLVEVNPAYSVPSSQVGGLEACIGCMQTRASVKLVKTCQEAATG ECQQCYCRPMWCLTCMGKWFASRQDPLRPDTWLASRVPCPTCRARFCILDVCTVR

Further analysis of the NOV105a protein yielded the following properties shown in Table 105B.

Table 105B. Protein Sequence Properties NOV105a	
PSort analysis:	0.6760 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Likely cleavage site between residues 29 and 30

A search of the NOV105a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 105C.

Table 105C. Geneseq Results for NOV105a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV105a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG81377	Human AFP protein sequence SEQ ID NO:272 - Homo sapiens, 362 aa. [WO200129221-A2, 26-APR-2001]	1..403 1..362	344/409 (84%) 345/409 (84%)	0.0

414

In a BLAST search of public sequence databases, the NOV105a protein was found to have homology to the proteins shown in the BLASTP data in Table 105D.

Table 105D. Public BLASTP Results for NOV105a				
Protein Accession Number	Protein/Organism/Length	NOV105a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC38627	SEQUENCE 271 FROM PATENT WO0129221 - Homo sapiens (Human), 362 aa.	1..403 1..362	345/409 (84%) 346/409 (84%)	0.0
Q9DCF3	0610039G24RIK PROTEIN - Mus musculus (Mouse), 362 aa.	1..403 1..362	311/403 (77%) 328/403 (81%)	e-176
Q96GP5	SIMILAR TO RIKEN CDNA 0610039G24 GENE - Homo sapiens (Human), 232 aa.	1..265 1..226	211/271 (77%) 212/271 (77%)	e-109
Q9VN16	CG14646 PROTEIN - Drosophila melanogaster (Fruit fly), 409 aa.	1..399 1..383	123/409 (30%) 202/409 (49%)	1e-55
Q95TM4	LD39811P - Drosophila melanogaster (Fruit fly), 393 aa.	20..399 4..367	117/390 (30%) 192/390 (49%)	1e-51

PFam analysis predicts that the NOV105a protein contains the domains shown in the Table 105E.

5

Table 105E. Domain Analysis of NOV105a			
Pfam Domain	NOV105a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 106.

The NOV106 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 106A.

Table 106A. NOV106 Sequence Analysis		
	SEQ ID NO: 301	1136 bp
NOV106a, CG57466-01 DNA Sequence	TTTCTGCAATGGGAGCCTCCGCCACCCACCCTGGAGCCACAGAAGGCCCAGAAGCCAA ATGGACAGCTGGTGAACCCCAACAACTTCTGGAAGAACCCGAAAGATGTGTGCGCCCA CGCCCATGGCCTCTCAGGGCCCAGGCCTGGGACGTGACCACCACTAACTGCTCAGCCA ATATCAACTTGACCCACCAGCCCTGGTTCCAGGTCCTGGAGCCGCAGTTCCGGCAGTT TCTCTTCTACCGCCACTGCCGCTACTTCCCCATGCTGCTGAACCACCCGGAGAAGTGC AGGGGCGATGTCTACCTGCTGGTGGTTGTCAAGTCGGTCATCACGCAGCACGACCGCC	

415

	GCGAGGCCATCCGCCAGACCTGGGCGCGAGCGGCAGTCCGCGGGTGGGGGCCGAGCGC CGTGCGCACCCTCTTCCTGCTGGGCACGGCCTCCAAGCAGGAGGAGCGCACGCACTAC CAGCAGCTGCTGGCCTACGAAGACGCCCTCTACGGCGACATCCTGCAGTGGGGCTTTC TCGACACCTTCTTCAACCTGACCCTCAAGGAGATCCACTTCCTCAAGTGGCTGGACAT CTACTGCCCCCACGTCCCCTTCATTTTCAAAGGCGACGATGACGTCTTCGTCAACCCC ACCAACCTGCTAGAATTTCTGGCTGACCGGCAGCCACAGGAAAACCTGTTCGTGGGCG ATGTCCTGCAGCACGCTCGGCCCATTCGCAGGAAAGACAACAAATACTACATCCCGGG GGCCCTGTACGGCAAGGCCAGCTATCCGCCGTATGCAGGCGGCGGTGGCTTCCTCATG GCCGGCAGCCTGGCCCGGCGCCTGCACCATGCCTGCGACACCCTGGAGCTCTACCCGA TCGACGACGTCTTTCTGGGCATGTGCCTGGAGGTGCTGGGCGTGCAGCCCACGGCCCA CGAGGGCTTCAAGACTTTCGGCATCTCCCGGAACCGCAACAGCCGCATGAACAAGGAG CCGTGCTTTTTCCGCGCCATGCTCGTGGTGCACAAGCTGCTGCCCCCTGAGCTGCTCG CCATGTGGGGGCTGGTGCACAGCAATCTCACCTGCTCCCGCAAGCTCCAGGTGCTCTG ACCCCAGCCGGGCTACTAGGACAGGCCAGGGCAC
	ORF Start: ATG at 9	ORF Stop: TGA at 1101
	SEQ ID NO: 302	364 aa	MW at 41853.8kD
NOV106a, CG57466-01 Protein Sequence	MGASATHPGATEGPEAKWTAGEPQQLLEEPERCVRPRPWPLRAQAWDVTTTNCSANIN LTHQPWFQVLEPQFRQFLFYRHCRYFPMLLNHPEKCRGDVYLLVVVKSVITQHDRREA IRQTWARAAVRGWGPSAVRTLFLLGTASKQEERTHYQQLLAYEDALYGDILQWGFLDT FFNLTLKEIHFLKWLDIYCPHVPFIFKGDDDVFVNPTNLLEFLADRQPQENLFVGDVL QHARPIRRKDNKYYIPGALYGKASYPPYAGGGGFLMAGSLARRLHHACDTLELYPIDD VFLGMCLEVLGVQPTAHEGFKTFGISRNRNSRMNKEPCFFRAMLVVHKLLPPELLAMW GLVHSNLTCSRKLQVL

Further analysis of the NOV106a protein yielded the following properties shown in Table 106B.

Table 106B. Protein Sequence Properties NOV106a	
PSort analysis:	0.6400 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.3122 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV106a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 106C.

Table 106C. Geneseq Results for NOV106a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV106a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB24035	Human PRO4397 protein sequence SEQ ID NO:42 - Homo sapiens, 402 aa. [WO200053750-A1, 14-SEP-2000]	72..352 84..380	149/300 (49%) 191/300 (63%)	4e-76
AAU29167	Human PRO polypeptide sequence #144 - Homo sapiens, 372 aa. [WO200168848-A2, 20-SEP-2001]	26..363 27..371	149/348 (42%) 207/348 (58%)	9e-76

416

AAB88404	Human membrane or secretory protein clone PSEC0159 - Homo sapiens, 372 aa. [EP1067182-A2, 10-JAN-2001]	26..363 27..371	149/348 (42%) 207/348 (58%)	9e-76
AAB49750	Human beta 1,3-N-acetylglucosamine transferase protein G4 - Homo sapiens, 372 aa. [WO200100848-A1, 04-JAN-2001]	26..363 27..371	149/348 (42%) 207/348 (58%)	9e-76
AAB49749	Human beta 1,3-N-acetylglucosamine transferase protein G4 - Homo sapiens, 372 aa. [WO200100848-A1, 04-JAN-2001]	26..363 27..371	149/348 (42%) 207/348 (58%)	9e-76

In a BLAST search of public sequence databases, the NOV106a protein was found to have homology to the proteins shown in the BLASTP data in Table 106D.

Table 106D. Public BLASTP Results for NOV106a				
Protein Accession Number	Protein/Organism/Length	NOV106a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAL32295	BETA-3-GALACTOSYLTRANSFERASE - Brachydanio rerio (Zebrafish) (Zebra danio), 418 aa.	46..364 101..417	199/319 (62%) 249/319 (77%)	e-121
AAL32297	BETA-3-GALACTOSYLTRANSFERASE - Brachydanio rerio (Zebrafish) (Zebra danio), 412 aa.	29..360 82..409	180/337 (53%) 244/337 (71%)	e-104
Q96EK0	UNKNOWN (PROTEIN FOR MGC:20513) - Homo sapiens (Human), 377 aa.	60..352 46..355	152/313 (48%) 198/313 (62%)	9e-76
CAC39768	SEQUENCE 175 FROM PATENT EP1067182 - Homo sapiens (Human), 372 aa.	26..363 27..371	149/348 (42%) 207/348 (58%)	3e-75
Q9C0J2	BETA-1,3-N-ACETYLGLUCOSAMINYLTRANSFERASE BGNT-3 - Homo sapiens (Human), 372 aa.	26..363 27..371	149/348 (42%) 207/348 (58%)	3e-75

PFam analysis predicts that the NOV106a protein contains the domains shown in the Table 106E.

<table>
<tr><td colspan="4" align="center">Table 106E. Domain Analysis of NOV106a</td></tr>
<tr><td align="center">Pfam Domain</td><td align="center">NOV106a Match Region</td><td align="center">Identities/ Similarities for the Matched Region</td><td align="center">Expect Value</td></tr>
<tr><td>PI3_PI4_kinase: domain 1 of 1</td><td>195..205</td><td>8/12 (67%)
10/12 (83%)</td><td>8.5</td></tr>
<tr><td>Galactosyl_T: domain 1 of 1</td><td>112..308</td><td>69/212 (33%)
148/212 (70%)</td><td>7.7e-45</td></tr>
</table>

Example 107.

The NOV107 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 107A.

<table>
<tr><td colspan="3" align="center">Table 107A. NOV107 Sequence Analysis</td></tr>
<tr><td></td><td>SEQ ID NO: 303</td><td>4091 bp</td></tr>
<tr><td>NOV107a,
CG57468-01 DNA Sequence</td><td colspan="2">AAGCAAGAGGCTGAGATGGATCTTGAGGCGGCAAAGAACGGAACAGCCTGGCGCCCCA
CGAGCGCGGAGGGCGACTTTGAACTGGGCATCAGCAGCAAACAAAAAAGGAAAAAAAC
GAAGACAGTGAAAATGATTGGAGTATTAACATTGTTTCGATACTCCGATTGGCAGGAT
AAATTGTTTATGTCGCTGGGTACCATCATGGCCATAGCTCACGGATCAGGTCTCCCCC
TCATGATGATAGTATTTGGAGAGATGACTGACAAATTTGTTGATACTGCAGGAAACTT
CTCCTTTCCAGTGAACTTTTCCTTGTCGCTGCTAAATCCAGGCAAAATTCTGGAAGAA
GAAATGACTAGATATGCATATTACTACTCAGGATTGGGTGCTGGAGTTCTTGTTGCTG
CCTATATACAAGTTTCATTTTGGACTTTGGCAGCTGGTCGACAGATCAGGAAAATTAG
GCAGAAGTTTTTTCATGCTATTCTACGACAGGAAATAGGATGGTTTGACATCAATGAC
ACCACTGAACTCAATACGCGGCTAACAGATGACATCTCCAAAATCAGTGAAGGAATTG
GTGACAAGGTTGGAATGTTCTTTCAAGCAGTAGCCACGTTTTTTGCAGGATTCATAGT
GGGATTCATCAGAGGATGGAAGCTCACCCTTGTGATAATGGCCATCAGCCCTATTCTA
GGACTCTCTGCAGCCGTTTGGGCAAAGATACTCTCGGCATTTAGTGACAAAGAACTAG
CTGCTTATGCAAAAGCAGGCGCCGTGGCAGAAGAGGCTCTGGGGGCCATCAGGACTGT
GATAGCTTTCGGGGGCCAGAACAAAGAGCTGGAAAGGTATCAGAAACATTTAGAAAAT
GCCAAAGAGATTGGAATTAAAAAAGCTATTTCAGCAAACATTTCCATGGGTATTGCCT
TCCTGTTAATATATGCATCATATGCACTGGCCTTCTGGTATGGATCCACTCTAGTCAT
ATCAAAAGAATATACTATTGGAAATGCAATGACAGTTTTTTTTTCAATCCTAATTGGA
GCTATGGCCATCGGAGAAACGCTCGTTTTGGCTCCTGAATATTCCAAAGCCAAATCGG
GGGCTGCGCATCTGTTTGCCTTGTTGGAAAAGAAACCAAATATAGACAGCCGCCAGTCA
AGAAGGGAAAAAGCCAGTAAGCGACACATGTGAAGGGAATTTAGAGTTTCGAGAAGTC
TCTTTCTTCTATCCATGTCGCCCAGATGTTTTCATCCTCCGTGGCTTATCCCTCAGTA
TTGAGCGAGGAAAGACAGTAGCATTTGTGGGGAGCAGCGGCTGTGGGAAAAGCACTTC
TGTTCAACTTCTGCAGAGACTTTATGACCCCGTGCAAGGACAAGTGGATGGTGTGGAT
GCAAAAGAATTGAATGTACAGTGGCTCCGTTCCCAAATAGCAATCGTTCCTCAAGAGC
CTGTGCTCTTCAACTGCAGCATTGCTGAGAACATCGCCTATGGTGACAACAGCCGTGT
GGTGCCATTAGATGAGATCAAAGAAGCCGCAAATGCAGCAAATATCCATTCTTTTATT
GAAGGTCTCCCTGAGAAATACAACACACAAGTTGGACTGAAAGGAGCACAGCTTTCTG
GCGGCCAGAAACAAAGACTAGCTATTGCAAGGGCTCTTCTCCAAAAACCCAAAATTTT
ATTGTTGGATGAGGCCACTTCAGCCCTCGATAATGACAGTGAGTGGCAGGTGGTTCAG
CATGCCCTTGATAAAGCCAGGACGGGAAGGACATGCCTAGTGGTCACTCACAGGCTCT
CTGCAATTCAGAACGCAGATTTGATAGTGGTTCTGCACAATGGAAAGATAAAGGAACA
AGGAACTCATCAAGAGCTCCTGAGAAATCGAGACATATATTTTAAGTTAGTGAATGCA
CAGTCAGCGAGCAAAGGTCGGACTACAATCGTGGTAGCACACCGACTTTCTACTATTC
GAAGTGCAGATTTGATTGTGACCCTAAAGGATGGAATGCTGGCGGAGAAAGGAGCACA
TGCTGAACTAATGCCAAAACGAGGTCTATATTATTCACTTGTGATGTCACAGGTAATG
CTTATGGGGACTCTTTCAGACTGTGGTAATAGTCTTCCTGAAGTCTCTCTATTAAAAA
TTTTAAAGTTAAACAAGCCTGAATGGCCTTTTGTGGTTCTGGGGACATTGGCTTCTGT
TCTAAATGGAACTGTTCATCCAGTATTTTCCATCATCTTTGCAAAAATTATAACCGTA
ATGTTTGGAAATAATGATCTTTTGTTTTTCCTCAAAATTTTTTTATATTCATTCCTTT</td></tr>
</table>

418

	TGTTTTTCCTCAAACAAGGTTTCAGCGTAGATTTTTGTTTGTTTGCTTTTCAGGGATT ATTTTACGGCAGAGCAGGGGAAATTTTAACGATGAGATTAAGACACTTGGCCTTCAAA GCCATGTTATATCAGGATATTGCCTGGTTTGATGAAAAGGAAAACAGCACAGGAGGCT TGACAACAATATTAGCCATAGATATAGCACAAATTCAAGGAGCAACAGGTTCCAGGAT TGGCGTCTTAACACAAAATGCAACTAACATGGGACTTTCAGTTATCATTTCCTTTATA TATGGATGGGAGATGACATTCCTGATTCTGAGTATTGCTCCAGTACTTGCCGTGACAG GAATGATTGAAACCGCAGCAATGACTGGATTTGCCAACAAAGATAAGCAAGAACTTAA GCATGCTGGAAAGGTAAAGATAGCAACTGAAGCTTTGGAGAATATACGTACTATAGTG TCATTAACAAGGGAAAAAGCCTTCGAGCAAATGTATGAAGAGATGCTTCAGACTCAAC ACAGGAGAAATACCTCGAAGAAAGCACAGATTATTGGAAGCTGTTATGCATTCAGCCA TGCCTTTATATATTTTGCCTATGCGGCAGGGTTTCGATTTGGAGCCTATTTAATTCAA GCTGGACGAATGTCAAATGCTTTATCTTTTGATAGAGTTTTTACTGCAATTGCATATG GAGCTATGGCCATCGGAGAAACGCTCGTTTTGGCTCCTGAATATTCCAAAGCCAAATC GGGGGCTGCGCATCTGTTTGCCTTGTTGGAAAAGAAACCAAATATAGACAGCCGCAGT CAAGAAGGGAAAAAGCCACTTTCACAGGACACATGTGAAGGGAATTTAGAGTTTCGAG AAGTCTCTTTCTTCTATCCATGTCGCCCAGATGTTTTCATCCTCCGTGGCTTATCCCT CAGTATTGAGCGAGGAAAGACAGTAGCATTTGTGGGGAGCAGCGGCTGTGGGAAAAGC ACTTCTGTTCAACTTCTGCAGAGACTTTATGACCCCGTGCAAGGACAACAGCTGTTTG ATGGTGTGGATGCAAAAGAATTGAATGTACAGTGGCTCCGTTCCCAAATAGCAATCGT TCCTCAAGAGCCTGTGCTCTTCAACTGCAGCATTGCTGAGAACATCGCCTATGGTGAC AACAGCCGTGTGGTGCCATTAGATGAGATCAAAGAAGCCGCAAATGCAGCAAATATCC ATTCTTTTATTGAAGGTCTCCCTAAATACAACACACAAGTTGGACTGAAAGGAGCACA GCTTTCTGGCGGCCAGAAACAAAGACTAGCTATTGCAAGGGCTCTTCTCCAAAAACCC AAAATTTTATTGTTGGATGAGGCCACTTCAGCCCTCGATAATGACAGTGAGAAGGTAC AGGTGGTTCAGCATGCCCTTGATAAAGCCAGGACGGGAAGGACATGCCTAGTGGTCAC TCACAGGCTCTCTGCAATTCAGAACGCAGATTTGATAGTGGTTCTGCACAATGGAAAG ATAAAGGAACAAGGAACTCATCAAGAGCTCCTGAGAAATCGAGACATATATTTTAAGT TAGTGAATGCACAGTCAGCGAGCAAAGGTCGGACTACAATCGTGGTAGCACACCGACT TTCTACTATTCGAAGTGCAGATTTGATTGTGACCCTAAAGGATGGAATGCTGGCGGAG AAAGGAGCACATGCTGAACTAATGGCAAAACGAGGTCTATATTATTCACTTGTGATGT CACAGGTAATGCTTATG**TGA**CATAATGCTAT	
	ORF Start: ATG at 16	**ORF Stop: TGA at 4078**
	SEQ ID NO: 304	**1354 aa**	**MW at 149167.3kD**
NOV107a, CG57468-01 Protein Sequence	MDLEAAKNGTAWRPTSAEGDFELGISSKQKRKKTKTVKMIGVLTLFRYSDWQDKLFMS LGTIMAIAHGSGLPLMMIVFGEMTDKFVDTAGNFSFPVNFSLSLLNPGKILEEEMTRY AYYYSGLGAGVLVAAYIQVSFWTLAAGRQIRKIRQKFFHAILRQEIGWFDINDTTELN TRLTDDISKISEGIGDKVGMFFQAVATFFAGFIVGFIRGWKLTLVIMAISPILGLSAA VWAKILSAFSDKELAAYAKAGAVAEEALGAIRTVIAFGGQNKELERYQKHLENAKEIG IKKAISANISMGIAFLLIYASYALAFWYGSTLVISKEYTIGNAMTVFFSILIGAMAIG ETLVLAPEYSKAKSGAAHLFALLEKKPNIDSRSQEGKKPVSDTCEGNLEFREVSFFYP CRPDVFILRGLSLSIERGKTVAFVGSSGCGKSTSVQLLQRLYDPVQGQVDGVDAKELN VQWLRSQIAIVPQEPVLFNCSIAENIAYGDNSRVVPLDEIKEAANAANIHSFIEGLPE KYNTQVGLKGAQLSGGQKQRLAIARALLQKPKILLLDEATSALDNDSEWQVVQHALDK ARTGRTCLVVTHRLSAIQNADLIVVLHNGKIKEQGTHQELLRNRDIYFKLVNAQSASK GRTTIVVAHRLSTIRSADLIVTLKDGMLAEKGAHAELMAKRGLYYSLVMSQVMLMGTL SDCGNSLPEVSLLKILKLNKPEWPFVVLGTLASVLNGTVHPVFSIIFAKIITVMFGNN DLLFFLKIFLYSFLLFFLKQGFSVDFCLFAFQGLFYGRAGEILTMRLRHLAFKAMLYQ DIAWFDEKENSTGGLTTILAIDIAQIQGATGSRIGVLTQNATNMGLSVIISFIYGWEM TFLILSIAPVLAVTGMIETAAMTGFANKDKQELKHAGKVKIATEALENIRTIVSLTRE KAFEQMYEEMLQTQHRRNTSKKAQIIGSCYAFSHAFIYFAYAAGFRFGAYLIQAGRMS NALSFDRVFTAIAYGAMAIGETLVLAPEYSKAKSGAAHLFALLEKKPNIDSRSQEGKK PLSQDTCEGNLEFREVSFFYPCRPDVFILRGLSLSIERGKTVAFVGSSGCGKSTSVQL LQRLYDPVQGQQLFDGVDAKELNVQWLRSQIAIVPQEPVLFNCSIAENIAYGDNSRVV PLDEIKEAANAANIHSFIEGLPKYNTQVGLKGAQLSGGQKQRLAIARALLQKPKILLL DEATSALDNDSEKVQVVQHALDKARTGRTCLVVTHRLSAIQNADLIVVLHNGKIKEQG THQELLRNRDIYFKLVNAQSASKGRTTIVVAHRLSTIRSADLIVTLKDGMLAEKGAHA ELMAKRGLYYSLVMSQVMLM

Further analysis of the NOV107a protein yielded the following properties shown in Table 107B.

Table 107B. Protein Sequence Properties NOV107a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)

419

SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV107a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 107C.

Table 107C. Geneseq Results for NOV107a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV107a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB81064	Cynomologous monkey P-glycoprotein variant 1 - Macaca fascicularis, 1280 aa. [WO200123565-A1, 05-APR-2001]	1..1299 1..1278	750/1312 (57%) 964/1312 (73%)	0.0
AAB81065	Cynomologous monkey P-glycoprotein variant 2 - Macaca fascicularis, 1283 aa. [WO200123565-A1, 05-APR-2001]	1..1299 1..1281	749/1312 (57%) 967/1312 (73%)	0.0
AAB81959	Human MDR1 - Homo sapiens, 1280 aa. [WO200121762-A2, 29-MAR-2001]	1..1299 1..1278	749/1324 (56%) 967/1324 (72%)	0.0
AAY58186	Human wild-type multidrug resistance-1 (MDR-1) protein - Homo sapiens, 1280 aa. [WO9961589-A2, 02-DEC-1999]	1..1299 1..1278	749/1324 (56%) 967/1324 (72%)	0.0
AAW44073	Human multidrug resistance P-glycoprotein MDR1 - Homo sapiens, 1280 aa. [WO9740160-A1, 30-OCT-1997]	1..1299 1..1278	749/1324 (56%) 967/1324 (72%)	0.0

5   In a BLAST search of public sequence databases, the NOV107a protein was found to have homology to the proteins shown in the BLASTP data in Table 107D.

420

Table 107D. Public BLASTP Results for NOV107a				
Protein Accession Number	Protein/Organism/Length	NOV107a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P23174	Multidrug resistance protein 3 (P-glycoprotein 3) - Cricetulus griseus (Chinese hamster), 1281 aa.	1..1299 1..1279	818/1303 (62%) 999/1303 (75%)	0.0
P21440	Multidrug resistance protein 2 (P-glycoprotein 2) - Mus musculus (Mouse), 1276 aa.	1..1299 1..1274	823/1306 (63%) 998/1306 (76%)	0.0
Q08201	Multidrug resistance protein 2 (P-glycoprotein 2) - Rattus norvegicus (Rat), 1278 aa.	1..1299 1..1276	823/1309 (62%) 999/1309 (75%)	0.0
CAC37764	SEQUENCE 1 FROM PATENT WO0123565 - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 1280 aa.	1..1299 1..1278	750/1312 (57%) 964/1312 (73%)	0.0
CAC37765	SEQUENCE 3 FROM PATENT WO0123565 - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 1283 aa.	1..1299 1..1281	749/1312 (57%) 967/1312 (73%)	0.0

PFam analysis predicts that the NOV107a protein contains the domains shown in the Table 107E.

Table 107E. Domain Analysis of NOV107a			
Pfam Domain	NOV107a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ABC_membrane: domain 1 of 2	57..350	115/301 (38%) 252/301 (84%)	3.3e-83
MVIN: domain 1 of 1	57..447	70/531 (13%) 263/531 (50%)	5.8
SAA_proteins: domain 1 of 1	518..524	6/7 (86%) 7/7 (100%)	3
ABC_tran: domain 1 of 2	424..609	76/199 (38%) 150/199 (75%)	3.1e-56

DsbD: domain 1 of 1	722..926	39/249 (16%) 126/249 (51%)	9.6
ABC_membrane: domain 2 of 2	722..1008	80/297 (27%) 222/297 (75%)	2.2e-43
ABC_tran: domain 2 of 2	1083..1270	77/202 (38%) 154/202 (76%)	7.1e-54
GidB: domain 1 of 1	1170..1312	29/202 (14%) 97/202 (48%)	6.6

Example 108.

The NOV108 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 108A.

Table 108A. NOV108 Sequence Analysis			
	SEQ ID NO: 305	520 bp	
NOV108a, CG59609-01 DNA Sequence	CCCCGTTCTATCAGCCATGGTCAACCCCACCAGGTTCTTAGACATCATCGTGGATGGT GAGCTCTTGGGACGTGTCTCCTTTGAGCTGTTTGCAGACAAGATTCCAAAGACAGCAG AAAATTTTTGTGCTCTAATCATTGGAGAGAAAGGATTTGGTTATAAAGGTTCCTACTT TCACAGAATTGTTCCTGGGTTTATGTGTCAGGGTGGTGACTTCACACAGCATAATGGC ACTGGTGGCAAGTCCATCTACGGGAAGAAATTTGATGATGAGAACTTCGTCCTAAATT ATACAGGTCCTGGCATCTTGTCCGTGGAGAATGCTGGACCCAACACAAATGGTTCCCA GTTTTTCATCTGCACTGCCATGTCTGAGTGGTTGGATGGCATGCAGGTGGTCTTTGGC AAGGGAAGGAAGGTGAGTATTGTGGAAGCCATGGAGTGCTTTGGGTCCACAAATGGCA AGACCAGCAAGAAGATCACCATTGCTGACTGTGGACAACTCTAATAGGTTTGACTT		
	ORF Start: ATG at 17	ORF Stop: TAA at 506	
	SEQ ID NO: 306	163 aa	MW at 17734.1kD
NOV108a, CG59609-01 Protein Sequence	MVNPTRFLDIIVDGELLGRVSFELFADKIPKTAENFCALIIGEKGFGYKGSYFHRIVP GFMCQGGDFTQHNGTGGKSIYGKKFDDENFVLNYTGPGILSVENAGPNTNGSQFFICT AMSEWLDGMQVVFGKGRKVSIVEAMECFGSTNGKTSKKITIADCGQL		

Further analysis of the NOV108a protein yielded the following properties shown in Table 108B.

Table 108B. Protein Sequence Properties NOV108a	
PSort analysis:	0.6400 probability located in microbody (peroxisome); 0.6000 probability located in plasma membrane; 0.4500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV108a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 108C.

422

Table 108C. Geneseq Results for NOV108a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV108a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU01195	Human cyclophilin A protein - Homo sapiens, 165 aa. [WO200132876-A2, 10-MAY-2001]	1..163 1..164	134/164 (81%) 147/164 (88%)	2e-75
AAW56028	Calcineurin protein - Mammalia, 165 aa. [WO9808956-A2, 05-MAR-1998]	1..163 1..164	134/164 (81%) 147/164 (88%)	2e-75
AAR13726	Bovine cyclophilin - Bos taurus, 163 aa. [US5047512-A, 10-SEP-1991]	2..163 1..163	133/163 (81%) 146/163 (88%)	5e-75
AAG65275	Haematopoietic stem cell proliferation agent related human protein #2 - Homo sapiens, 164 aa. [JP2001163798-A, 19-JUN-2001]	2..163 1..163	133/163 (81%) 146/163 (88%)	9e-75
AAP90431	Cyclophilin - Homo sapiens (human), 164 aa. [EP326067-A, 02-AUG-1989]	2..163 1..163	133/163 (81%) 146/163 (88%)	9e-75

In a BLAST search of public sequence databases, the NOV108a protein was found to have homology to the proteins shown in the BLASTP data in Table 108D.

Table 108D. Public BLASTP Results for NOV108a				
Protein Accession Number	Protein/Organism/Length	NOV108a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC39529	SEQUENCE 26 FROM PATENT WO0132876 - Homo sapiens (Human), 165 aa.	1..163 1..164	134/164 (81%) 147/164 (88%)	8e-75
Q9BRU4	PEPTIDYLPROLYL ISOMERASE A (CYCLOPHILIN A) - Homo sapiens (Human), 165 aa.	1..163 1..164	134/164 (81%) 146/164 (88%)	2e-74
P04374	Peptidyl-prolyl cis-trans isomerase A (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin A) (Cyclosporin A-binding protein) - Bos taurus (Bovine), and, 163 aa.	2..163 1..163	133/163 (81%) 146/163 (88%)	2e-74

| P05092 | Peptidyl-prolyl cis-trans isomerase A (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin A) (Cyclosporin A-binding protein) - Homo sapiens (Human),, 164 aa. | 2..163
1..163 | 133/163 (81%)
146/163 (88%) | 3e-74 |
| Q9TTC6 | CYCLOPHILIN 18 - Oryctolagus cuniculus (Rabbit), 164 aa. | 1..163
1..164 | 133/164 (81%)
147/164 (89%) | 5e-74 |

PFam analysis predicts that the NOV108a protein contains the domains shown in the Table 108E.

Table 108E. Domain Analysis of NOV108a

Pfam Domain	NOV108a Match Region	Identities/ Similarities for the Matched Region	Expect Value
pro_isomerase: domain 1 of 1	5..163	101/181 (56%) 137/181 (76%)	5.2e-79

Example 109.

The NOV109 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 109A.

Table 109A. NOV109 Sequence Analysis

	SEQ ID NO: 307	887 bp
NOV109a, CG59613-01 DNA Sequence	GATATCATTTTTT**ATG**GCAGCCATTGTTAAGCCTCCAGAACCTATACCTTTAAAATGG TTAACAGATAAGCCAGTTTGGATAGAACAATGGCCACTGAGTAAAGAGAAACTGGAGG CTTTAGAGGATTTGGTTACTGAACAATTCTCAATAATCATTTTCCAAAAAGTGAACCT ACACAGCATGAAAGTATCACACATTTCCTTAGTGCAGCTAACCCTGTGTGACCAGGGC TTCAACACATACCACTGTGACCACAACCTAGCCATGAGCATGAGCCTCACCAGCATGT CCAAAATGCTAAAATACAACAATGGCAGTGAAGACATCACTACATGGAGGGCTGAAGG TACTATGGATCTCTTGGTGCTAGAATTTGAAGCACTAAATCAAGAGAACTTTGTGGAC TGTGAATTGAAGTTAATGACTCTAGATGTTGAGCAACTTGAAATTCCAGAACAAGAGT ACAGCTGTGTAATAAAGATGCATTCTAGTGAATTTGTTCATATATGCCAAGATCTCAG TCATATTGGAGAGTCTGCTATAATTTCTTGTGCAAAAGATGGAGTGAATTTTTCTGCA AATGGAGAACTTGGACATGGAAACATTGCCACAATTGCCCAAACAAGTAATTACAATA AAGAAGAGGAGGCTGTTGCCATAATGATGAATGGGCCAGTTCAGCTAACTTTTGCACT AAGTTACTTAAATTTCTTTATAACAGGCACTCCACTCTCTCAGATGCACCCCTTGCTG GAGAGTATAAGATTGCCGGATATGGAACATTTAAAGTATTATTTGGCTCCCAAAATTG AGGATGAAAAAGGATTT**TAG**AAATTCTTAGAATCCAAGAAAATAAAACTAAGCTCTTT GAAAATTGCTTCTGAGA
	ORF Start: ATG at 14	ORF Stop: TAG at 830
	SEQ ID NO: 308	272 aa	MW at 30831.1kD
NOV109a, CG59613-01 Protein Sequence	MAAIVKPPEPIPLKWLTDKPVWIEQWPLSKEKLEALEDLVTEQFSIIIFQKVNLHSMK VSHISLVQLTLCDQGFNTYHCDHNLAMSMSLTSMSKMLKYNNGSEDITTWRAEGTMDL LVLEFEALNQENFVDCELKLMTLDVEQLEIPEQEYSCVIKMHSSEFVHICQDLSHIGE SAIISCAKDGVNFSANGELGHGNIATIAQTSNYNKEEEAVAIMMNGPVQLTFALSYLN FFITGTPLSQMHPLLESIRLPDMEHLKYYLAPKIEDEKGF	

Further analysis of the NOV109a protein yielded the following properties shown in Table 109B.

Table 109B. Protein Sequence Properties NOV109a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Likely cleavage site between residues 19 and 20

A search of the NOV109a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 109C.

Table 109C. Geneseq Results for NOV109a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV109a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY51639	Human PCNA protein fragment - Homo sapiens, 261 aa. [WO200008164-A2, 17-FEB-2000]	25..271 8..260	158/255 (61%) 184/255 (71%)	8e-78
AAY52010	Human PCNA protein - Homo sapiens, 261 aa. [DE19840771-A1, 10-FEB-2000]	25..271 8..260	158/255 (61%) 184/255 (71%)	8e-78
AAB43712	Human cancer associated protein sequence SEQ ID NO:1157 - Homo sapiens, 269 aa. [WO200055350-A1, 21-SEP-2000]	25..271 16..268	158/255 (61%) 184/255 (71%)	8e-78
AAG75139	Human colon cancer antigen protein SEQ ID NO:5903 - Homo sapiens, 268 aa. [WO200122920-A2, 05-APR-2001]	25..269 16..266	157/253 (62%) 182/253 (71%)	5e-77
AAW90758	Human PCNA protein fragment #2 - Homo sapiens, 236 aa. [DE19840771-A1, 10-FEB-2000]	39..268 1..236	149/238 (62%) 171/238 (71%)	7e-73

In a BLAST search of public sequence databases, the NOV109a protein was found to have homology to the proteins shown in the BLASTP data in Table 109D.

425

Table 109D. Public BLASTP Results for NOV109a				
Protein Accession Number	Protein/Organism/Length	NOV109a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P12004	Proliferating cell nuclear antigen (PCNA) (Cyclin) - Homo sapiens (Human), 261 aa.	25..271 8..260	158/255 (61%) 184/255 (71%)	3e-77
P04961	Proliferating cell nuclear antigen (PCNA) (Cyclin) - Rattus norvegicus (Rat), 261 aa.	25..271 8..260	158/255 (61%) 185/255 (71%)	5e-77
P57761	Proliferating cell nuclear antigen (PCNA) - Cricetulus griseus (Chinese hamster), 261 aa.	25..271 8..260	158/255 (61%) 184/255 (71%)	7e-77
Q91ZH2	11 DAYS EMBRYO CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:2700095L20, FULL INSERT SEQUENCE - Mus musculus (Mouse), 261 aa.	25..272 8..261	156/256 (60%) 183/256 (70%)	1e-75
P17918	Proliferating cell nuclear antigen (PCNA) (Cyclin) - Mus musculus (Mouse), 261 aa.	25..270 8..259	155/254 (61%) 182/254 (71%)	5e-75

PFam analysis predicts that the NOV109a protein contains the domains shown in the Table 109E.

Table 109E. Domain Analysis of NOV109a			
Pfam Domain	NOV109a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PCNA: domain 1 of 1	23..143	46/128 (36%) 83/128 (65%)	2.3e-20
PCNA_C: domain 1 of 1	145..265	59/131 (45%) 98/131 (75%)	1.6e-45

Example 110.

The NOV110 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 110A.

Table 110A. NOV110 Sequence Analysis		
	SEQ ID NO: 309	1233 bp
NOV110a, CG59619-01 DNA Sequence	TGGCAATGGAAGAAGAGATCCCCGCGCTCTTCATTGACAATGGCTCCGGCATGTGGAA AGCAGCTTTGCTGGGAGACAATGCCCTCCGAGCCATATTCCCCTCCATCATCGGGCAC CCCCGGCACCAGGGCGTGATGGTGGGCATGGGCCAGAAGGACTCCTACGTGGGCGACC AGGCCCAGAGCAAGTGCGGCATCCTGACCCTGAAGTACCCCATCAAGCATGGCATCGT CACAAACTGGGACGACATGGAGAAGATCTGGCACCATGTTTTCTACAACGAGCTGTGC GTGGCCCTGGAGGAGCAGGTGGTGCTGCTGACCGAGGCCCCGCTAAACCCCAGGGCCA ATAGGGAGAAGATGACTCAGATCATGTTTAAGACCTTCAACACCCAGGCCATGTACGT GGCCATTCAGGCCGTGCTGACCCTCCACAGCTCTGGTTGCACCACTGGCATTGTCATG GACTCTGGAGATGGGGTCACCCACACAGTGCCCATCTACGAGCGCCACACCCTCCCTC ACACCATCTTGCATCTGGACCTGGCTGGCCAGGACCTTACTGACTACCTCATGAAGAT CCCTACCTACCGCAGCTATAGCTTCAACACCATGGCCAAGTGGAAAATCGTGCGCAAC ATCAAGGAGAAGCTATGCTATGTCGCTCTGGACTTCGAGGAGGAGATGGCCACTGCTG CATCCTCCTCCTCCCTGGAGAAGAGCTACGAGCTGCCTGACAGCCAGGCCATCATTAT TAGCAATGAGCGGTTCCGGTGTCCGGAGGCACTGTTCCAGCCTTCCTTCCTGGGCATG GAATCCTGTGGCATCCATGAAAGTACCTTCAACTCCATCATGAAGTGTGATATGGACA TCCCCAAAGACCTGTACGCCAACACAGTGCTGTCTGGCGTCACCACCATGTACCCTGG CATCCCCAATAGGATGCAGAAGGAGATCACTGCCCTGGCATCCAGCACCATGAAGATC AAGATATCGTGCCCCATCGTGCCCCCAGAGTGCAAGTACTTTGTGTGGATCGGTGGCT CCATCCTGGCCTCACTGTCCACCTTCCAGCAGATGTGGATTAGCAAGCAGGAGTACAA CGAGTCGGGCCCCTCCATCATCCACCGCAAATGGACTGCGAGCAGATGCATAGCATTT GCTGCATGGGTTAATTCAGAAGTATAAATTTGCCCCTGGCAAATGCATATACCTCATG CTAGCCTCACGATAC
	ORF Start: ATG at 6	ORF Stop: TAA at 1185
	SEQ ID NO: 310	393 aa MW at 44147.5kD
NOV110a, CG59619-01 Protein Sequence	MEEEIPALFIDNGSGMWKAALLGDNALRAIFPSIIGHPRHQGVMVGMGQKDSYVGDQA QSKCGILTLKYPIKHGIVTNWDDMEKIWHHVFYNELCVALEEQVVLLTEAPLNPRANR EKMTQIMFKTFNTQAMYVAIQAVLTLHSSGCTTGIVMDSGDGVTHTVPIYERHTLPHT ILHLDLAGQDLTDYLMKIPTYRSYSFNTMAKWKIVRNIKEKLCYVALDFEEEMATAAS SSSLEKSYELPDSQAIIISNERFRCPEALFQPSFLGMESCGIHESTFNSIMKCDMDIP KDLYANTVLSGVTTMYPGIPNRMQKEITALASSTMKIKISCPIVPPECKYFVWIGGSI LASLSTFQQMWISKQEYNESGPSIIHRKWTASRCIAFAAWVNSEV

Further analysis of the NOV110a protein yielded the following properties shown in Table 110B.

Table 110B. Protein Sequence Properties NOV110a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.1547 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV110a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 110C.

Table 110C. Geneseq Results for NOV110a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV110a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU32060	Novel human secreted protein #2551 - Homo sapiens, 399 aa. [WO200179449-A2, 25-OCT-2001]	1..376 25..397	315/376 (83%) 336/376 (88%)	e-180
AAB43991	Human cancer associated protein sequence SEQ ID NO:1436 - Homo sapiens, 413 aa. [WO200055350-A1, 21-SEP-2000]	1..376 39..411	311/376 (82%) 336/376 (88%)	e-179
AAP61532	Sequence of beta-actin - Homo sapiens, 375 aa. [EP174608-A, 19-MAR-1986]	1..376 1..373	311/376 (82%) 335/376 (88%)	e-179
AAB12985	Human beta-actin protein sequence - Homo sapiens, 374 aa. [US6087398-A, 11-JUL-2000]	2..376 1..372	310/375 (82%) 334/375 (88%)	e-178
AAR50328	Drug resistant structural protein - Homo sapiens, 375 aa. [JP06038773-A, 15-FEB-1994]	1..376 1..373	309/376 (82%) 335/376 (88%)	e-178

In a BLAST search of public sequence databases, the NOV110a protein was found to have homology to the proteins shown in the BLASTP data in Table 110D.

Table 110D. Public BLASTP Results for NOV110a				
Protein Accession Number	Protein/Organism/Length	NOV110a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P02571	Actin, cytoplasmic 2 (Gamma-actin) - Homo sapiens (Human),, 375 aa.	1..376 1..373	315/376 (83%) 336/376 (88%)	e-179
ATBOG	actin gamma (tentative sequence) - bovine, 374 aa.	2..376 1..372	314/375 (83%) 335/375 (88%)	e-179
P53505	Actin, cytoplasmic type 5 - Xenopus laevis (African clawed frog), 376 aa.	2..376 3..374	313/375 (83%) 335/375 (88%)	e-178
P29751	Actin, cytoplasmic 1 (Beta-actin) - Oryctolagus cuniculus (Rabbit), 375 aa.	1..376 1..373	311/376 (82%) 337/376 (88%)	e-178

428

O93400	ACTIN, CYTOPLASMIC 1 (BETA-ACTIN) (CYTOPLASMIC BETA ACTIN) - Xenopus laevis (African clawed frog), 375 aa.	1..376 1..373	311/376 (82%) 336/376 (88%)	e-178

PFam analysis predicts that the NOV110a protein contains the domains shown in the Table 110E.

Table 110E. Domain Analysis of NOV110a			
Pfam Domain	**NOV110a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
actin: domain 1 of 1	1..378	284/382 (74%) 336/382 (88%)	2.2e-227

Example 111.

The NOV111 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 111A.

Table 111A. NOV111 Sequence Analysis		
	SEQ ID NO: 311	1197 bp
NOV111a, CG59621-01 DNA Sequence	AACCATGTCTAAGCGGGAGTCCTTTAACCTGGAAAGTTATGAATTGGACAAAAGCTTC TGGCTAACCAGATTCACTGAACTGAAGGGCACAGGTTGCAAAGTGCCCCAAGATGTCT TGCAAAAATTGCTGGAATCTTTACAGGAGAACCACTTCCAAGAAGATGAGCAGTTTCT GGGAGCCGTTATGCCAAGGCTTCGCATTGGAATGGATACTTGTGCCATTTCTTTGAGG CATGGTGGGCTTTCCTTGGTTCAAACCACAGATTACATTTACCCGATCGTAGACGACC CTTACATGATGGGCAGGATAGCATGTGCCAATGTCCTCAGTGACCTCTATGCAATGGG GGTCACAGAATGTGACAATATGCTGATGCTCCTTGGAGTCAGTAATAAAATGACCGAC AGGGAAAGGGATAAAGTGATGCCTCTGATTATCCAAAGTTTTAAAGATGCAGCTGAGG AAGCAGGAATGTCTGTAATGGTCAGCCAAACAGTACTAAATCCCTGGATTGTCCTGGG AGGAGTCACTACCACTGTCTTCCAGCCCAATGAATTTATCATGCCAGACAATGCAGTG CCAGGGGACGTGCTGGTGTTGACAAAACCCCTGGGGACACAGGTGGCAGTGGCTGTGC ACCAGTGGCTGGATATTCCTTTGAAATGGAATAAGATTAAGCTAGTGGTCACCGAAGA TGTAGAGCTGGCCAACCAGGAGGCGATGATGAACATGGTGAGGCTCAACAGGACAGCT GCAGGACTCATGCACACGTTCAATGCCCACATGGCCACTGACATCACGGGCTTCGGGA TTTTGGGCCACGTGCAGAACCTAGCCAAGCAGCAGAGGAACGAGGTGTCGTTTGTAAT TCACAACCTCCTGGTGCTGGCCAAGATGGCTGCGGTGAGCAAGGCCTGCGGAAACATG TTCAGCCTCATGCATGGGACCTGCCCGGAGACCTCAGGCGGCCTTCTGATCTGTTTAC CATGTCAGCAAGCAGCTCGGTTCTGTGCAGAGATAAAGTCCCCCAAATATAGTGAAGG CCACCAAGCATGGATTATTGGGATTGTAGAGAAGGGCAACCACACAGCCAGAATCATA GACAAACCCCAGATCATCAAGGTTGCACCACAAGTGGCCACTCAAAATGTGAATCTCA CACCCGGGGCCACATCTTAATCTAGACAGAAATAGCT
	ORF Start: ATG at 5	ORF Stop: TAA at 1178
	SEQ ID NO: 312	391 aa	MW at 43193.9kD
NOV111a, CG59621-01 Protein Sequence	MSKRESFNLESYELDKSFWLTRFTELKGTGCKVPQDVLQKLLESLQENHFQEDEQFLG AVMPRLRIGMDTCAISLRHGGLSLVQTTDYIYPIVDDPYMMGRIACANVLSDLYAMGV TECDNMLMLLGVSNKMTDRERDKVMPLIIQSFKDAAEEAGMSVMVSQTVLNPWIVLGG VTTTVFQPNEFIMPDNAVPGDVLVLTKPLGTQVAVAVHQWLDIPLKWNKIKLVVTEDV ELANQEAMMNMVRLNRTAAGLMHTFNAHMATDITGFGILGHVQNLAKQQRNEVSFVIH NLLVLAKMAAVSKACGNMFSLMHGTCPETSGGLLICLPCQQAARFCAEIKSPKYSEGH QAWIIGIVEKGNHTARIIDKPQIIKVAPQVATQNVNLTPGATS

429

Further analysis of the NOV111a protein yielded the following properties shown in Table 111B.

Table 111B. Protein Sequence Properties NOV111a	
PSort analysis:	0.8500 probability located in endoplasmic reticulum (membrane); 0.4400 probability located in plasma membrane; 0.1000 probability located in mitochondrial inner membrane; 0.1000 probability located in Golgi body
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV111a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 111C.

Table 111C. Geneseq Results for NOV111a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV111a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB58174	Lung cancer associated polypeptide sequence SEQ ID 512 - Homo sapiens, 250 aa. [WO200055180-A2, 21-SEP-2000]	166..391 20..243	168/227 (74%) 189/227 (83%)	2e-88
AAO01161	Human polypeptide SEQ ID NO 15053 - Homo sapiens, 122 aa. [WO200164835-A2, 07-SEP-2001]	147..264 1..118	81/119 (68%) 92/119 (77%)	2e-36
AAB53700	Human colon cancer antigen protein sequence SEQ ID NO:1240 - Homo sapiens, 106 aa. [WO200055351-A1, 21-SEP-2000]	42..99 1..58	53/58 (91%) 54/58 (92%)	1e-24

In a BLAST search of public sequence databases, the NOV111a protein was found to have homology to the proteins shown in the BLASTP data in Table 111D.

430

Table 111D. Public BLASTP Results for NOV111a				
Protein Accession Number	**Protein/Organism/Length**	**NOV111a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9BVT4	SELENOPHOSPHATE SYNTHETASE , HUMAN SELENIUM DONOR PROTEIN - Homo sapiens (Human), 392 aa.	1..391 1..392	364/392 (92%) 367/392 (92%)	0.0
P49903	Selenide,water dikinase 1 (EC 2.7.9.3) (Selenophosphate synthetase 1) (Selenium donor protein 1) - Homo sapiens (Human), 383 aa.	1..375 1..376	348/376 (92%) 351/376 (92%)	0.0
AAC50958	SELENOPHOSPHATE SYNTHETASE 2 - Homo sapiens (Human), 448 aa.	2..391 33..441	272/411 (66%) 313/411 (75%)	e-147
Q99611	Selenide,water dikinase 2 (EC 2.7.9.3) (Selenophosphate synthetase 2) (Selenium donor protein 2) - Homo sapiens (Human), 448 aa.	2..391 33..441	272/411 (66%) 313/411 (75%)	e-147
AAC53024	SELENOPHOSPHATE SYNTHETASE 2 - Mus musculus (Mouse), 452 aa.	2..387 36..441	267/407 (65%) 307/407 (74%)	e-146

PFam analysis predicts that the NOV111a protein contains the domains shown in the Table 111E.

Table 111E. Domain Analysis of NOV111a			
Pfam Domain	**NOV111a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AIRS: domain 1 of 1	32..188	29/180 (16%) 113/180 (63%)	3e-18
AIRS_C: domain 1 of 1	191..367	34/197 (17%) 125/197 (63%)	1.1e-20

Example 112.

The NOV112 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 112A.

431

Table 112A. NOV112 Sequence Analysis		
	SEQ ID NO: 313	1544 bp
NOV112a, CG59625-01 DNA Sequence	CGATGGGACACAGACAGGTCACCCCAGCTCTGATCTTTGCCATCACAGTTGCTACAAT CGGCTCTTTCCAGTTTGGCTACAACACTGGGGTCATCAATGCTCCTGAGACGGTGCAG ATCATAAAGGAATTTATCAATAAAACTTTGACGGACAAGGCAAATGCCCCTCCCTCTG AGGTGCTGCTCACGAATCTCTGGTCCTTGTCTGTGGCCATATTTTCCGTCGGGGGTAT GATCGGCTCCTTTTCCGTCGGACTCTTTGTTAACCGCTTTGGCAGGAGGCGCAATTCA ATGCTGATTGTCAACCTGTTGGCTGCCACTGGTGGCTGCCTTATGGGACTGTGTAAAA TAGCTGAGTCAGTTGAAATGCTGATCCTGGGCCGCTTGGTTATTGGCCTCTTCTGCGG ACTCTGCACAGGTTTTGTGCCCATGTACATTGGAGAGATCTCGCCTACTGCCCTGAGG GGTGCCTTTGGCACTCTCAACCAGCTGGGCATAGTTATTGGAATTCTGGTGGCCCAGG TAATCTTTGGTCTGGAACTCATCCTTGGGTCTGAAGAGCTATGGCCGGTGCTATTAGG CTTTACCATCCTTCCAGCTATCCTGCAAAGTGCAGCCCTTCCATGTTGCCCTGAAAGT CCCAGATTTTTGCTCATTAACAGAAAAAAAGAGGAGAATGCTACGCGGGTCCTCCAGC GGTTGTGGGGCACCCAGGATGTATCCCAAGACATCCAGGAGATGAAAGATGAGAGTGC AAGGATGTCACAAGAAAAGCAAGTCACCGTGCTGGAGCTCTTTAGAGTGTCCAGCTAC CGACAGCCCATCATCATTTCCATTGTGCTCCAGCTCTCTCAGCAGCTCTCTGGGATCA ATGCTGTGGTGTTCTATTACTCAACAGGAATCTTCAAGGATGCAGGTGTTCAACAGCC CATCTATGCCACCATCAGCGCGGGTGTGGTTAATACTATCTTCACTTTACTTTCTGTA GTAGCTCAGATGCTGTTTTCATGGAAAGGAAAACTGAAGTTTCATGTCATAACTGTTT CTTTGTTATTAAAGCTGGGTTACACTGTCTTTAAATTTAATCTTCTGTGTTCCTTCCT CTTACAGAATCACTATAATGGGATGAGCTTTGTCTGTATTGGGGCTATCTTGGTCTTT GTGGCCTGTTTTGAAATTGGACCAGGCCCCATTCCCTGGTTTATTGTGGCCGAACTCT TCAGCCAGGGCCCCCGCCCAGCTGCGATGGCAGTGGCCGGCTGCTCCAACTGGACCTC CAACTTCCTAGTCGGATTGCTCTTCCCCTCTGCTGCTTACTATTTAGGAGCCTACGTT TTTATTATCTTCACCGGCTTCCTCATTACCTTCTTGGCCTTTACCTTCTTCAAAGTCC CTGAGACCCGTGGCAGGACTTTTGAGGATATCACACGGGCCTTTGAAGGGCAGGCACA CGGTGCAGATAGATCTGGGAAGGACGGCGTCATGGGGATGAACAGCATCGAGCCTGCT AAGGAGACCACCACCAATGTCTAAGTCATGCCTCCT
	ORF Start: ATG at 3	ORF Stop: TAA at 1530
	SEQ ID NO: 314	509 aa	MW at 55571.7kD
NOV112a, CG59625-01 Protein Sequence	MGHRQVTPALIFAITVATIGSFQFGYNTGVINAPETVQIIKEFINKTLTDKANAPPSE VLLTNLWSLSVAIFSVGGMIGSFSVGLFVNRFGRRRNSMLIVNLLAATGGCLMGLCKI AESVEMLILGRLVIGLFCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQV IFGLELILGSEELWPVLLGFTILPAILQSAALPCCPESPRFLLINRKKEENATRVLQR LWGTQDVSQDIQEMKDESARMSQEKQVTVLELFRVSSYRQPIIISIVLQLSQQLSGIN AVVFYYSTGIFKDAGVQQPIYATISAGVVNTIFTLLSVVAQMLFSWKGKLKFHVITVS LLLKLGYTVFKFNLLCSFLLQNHYNGMSFVCIGAILVFVACFEIGPGPIPWFIVAELF SQGPRPAAMAVAGCSNWTSNFLVGLLFPSAAYYLGAYVFIIFTGFLITFLAFTFFKVP ETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPAKETTTNV	

Further analysis of the NOV112a protein yielded the following properties shown in Table 112B.

Table 112B. Protein Sequence Properties NOV112a	
Psort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 22 and 23

A search of the NOV112a protein against the Geneseq database, a proprietary

database that contains sequences published in patents and patent publication, yielded

several homologous proteins shown in Table 112C.

432

Table 112C. Geneseq Results for NOV112a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV112a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY27289	Glucose transporter protein GLUT3 - Homo sapiens, 494 aa. [US5942398-A, 24-AUG-1999]	1..505 1..492	389/505 (77%) 431/505 (85%)	0.0
AAR11360	Glucose Transporter Protein from CHO cells - Cricetulus sp, 492 aa. [WO9103554-A, 21-MAR-1991]	4..491 6..481	289/489 (59%) 364/489 (74%)	e-156
AAW17835	Human glucose transporter GLUT-1 - Homo sapiens, 492 aa. [WO9715668-A2, 01-MAY-1997]	4..491 6..481	287/489 (58%) 362/489 (73%)	e-155
AAW93000	Human GLUT1 protein - Homo sapiens, 492 aa. [WO9618957-A1, 20-JUN-1996]	4..491 6..481	284/489 (58%) 360/489 (73%)	e-153
AAB30522	Amino acid sequence of a consensus GLUT polypeptide - Synthetic, 493 aa. [US6136547-A, 24-OCT-2000]	6..501 10..490	289/496 (58%) 357/496 (71%)	e-151

In a BLAST search of public sequence databases, the NOV112a protein was found to have homology to the proteins shown in the BLASTP data in Table 112D.

Table 112D. Public BLASTP Results for NOV112a				
Protein Accession Number	Protein/Organism/Length	NOV112a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P11169	Solute carrier family 2, facilitated glucose transporter, member 3 (Glucose transporter type 3, brain) - Homo sapiens (Human), 496 aa.	1..509 1..496	446/510 (87%) 468/510 (91%)	0.0
P47842	Solute carrier family 2, facilitated glucose transporter, member 3 (Glucose transporter type 3, brain) - Canis familiaris (Dog), 495 aa.	1..507 1..494	400/507 (78%) 446/507 (87%)	0.0
P47843	Solute carrier family 2, facilitated glucose transporter, member 3 (Glucose transporter type 3, brain) - Ovis aries (Sheep), 494 aa.	1..505 1..492	389/505 (77%) 431/505 (85%)	0.0

433

| P58352 | Solute carrier family 2, facilitated glucose transporter, member 3 (Glucose transporter type 3, brain) - Bos taurus (Bovine), 494 aa. | 1..505
1..492 | 390/505 (77%)
431/505 (85%) | 0.0 |
| Q07647 | Solute carrier family 2, facilitated glucose transporter, member 3 (Glucose transporter type 3, brain) - Rattus norvegicus (Rat), 493 aa. | 1..508
1..492 | 380/508 (74%)
422/508 (82%) | 0.0 |

PFam analysis predicts that the NOV112a protein contains the domains shown in the Table 112E.

Table 112E. Domain Analysis of NOV112a			
Pfam Domain	NOV112a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Herpes_glycop: domain 1 of 1	1..249	40/417 (10%) 171/417 (41%)	7.2
GntP_permease: domain 1 of 1	65..329	70/478 (15%) 185/478 (39%)	2.5
sugar_tr: domain 1 of 1	12..478	188/503 (37%) 410/503 (82%)	2.2e-158

Example 113.

5   The NOV113 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 113A.

Table 113A. NOV113 Sequence Analysis		
	SEQ ID NO: 315	1731 bp
NOV113a, CG59887-01 DNA Sequence	ACTACTTCGCCGACACTCGCCAGCCTCGGCTACGAGCAAAAAATGCACCGCACCATGA GCTCGTTCACCTCGTTTGCCCTGGCCTTTTCCATGGTCTCGATCAACACCGGCGTGGT CACGCTGTTCGCCGACCCGTTCAACCGCGTCGGGGGCATCGGCATCCTCCTGTGGCTG TTGGTGATCCCGCTGGTGTGCTGCATCGTCATGGTCTACTGCCACCTGGCCGGGCGCA TTCCGCTCACCGGCTACGCCTACCAATGGTCCAGCCGATTGGCGGGCAATCACTTCGG CTGGTTTACCGGCTGGGTGGCGTTCACCTCGTTTGTCGCCGGTACAGCCGCCACCTCG GCGGCCATCGGTACGGTGTTCGCCACCGGAGATCTGGGCCAACCCGACACAGGGTCAGA TCCAGGGCCTGAGCATCGGCGCCACGCTGGTGGTGGGCTTGCTGAATATCTGCGGGAT TCGCCTGGCCACCCGGATCAACGACATCGGCGCGATCATCGAAATCATCGGCACGGTA CTGCTGGCGATTGCGTTGTTCTTCGGGGTGTTTTTCTTCTTTGAGCACACCCAGGGCG TGGCGATCCTGACCTCCGCGCAACCAGTGAGCGGCGGCACGCTCAGCTTCACCACCAT CGCCCTCGCCACCTTGCTGCCGGTCTCGGTGCTGCTGGGTTGGGAAGGCGCCGCCGAC CTGTCCGAGGAAACCAAGGACCCACGCCGCGCCGCGCCCCGGGCGATGATTCGTGCGG TGCTGGTGTCCAGCGTATTGGGCTTCGTGGTGTTCGCCTTGCTGAGCATCGCGATCCC GGGCTCGGTCAGCGAACTGCTCAGCCACAGCGAAAACCCGGTGATCAATATCGTGCGC CTGCAACTGGGCAATGCCGCCGGCGTGGGCATGATCGTGATCGCTTTCGCCTCGATCC TCGCCTGCCTGATCGCCAACATGGCGGTGGCCACGCGCATGACCTTCGCCCTGTCCCG GGACAACATGCTGCCGGGGCTCCAAGGTGCTGGCGAAGATCAACCCGCACTTCGGCACG CCGGTCGCCGCCATCGTGCTGATCACCGCCATCGCCGTGCTGCTGAACCTGGCGAGTG

434

	GCGGGTTTGTCACGGCGATCTACTCGATGGTCGGCCTGACCTACTACTGCACTTACCT GCTGACGCTGATTGCCGCGTACCTGGCCTATAAAAACGGCCGGATGCCGGGGGCGCCT GCGGGCGTGTTCAGCCTGGGCCGCTGGTTGCTGCCGATGATTATCCTCGGCGGCCTGT GGGCCATCGCGGTGATCCTGACCCTGAGCGTGCCGGAAGAAAGCCACACTGGCGCTAT CACCACCGGGGTTACACTCGGCGTGGGCGTGTTGTGGTGGTTGTTTTCACTGCGCACG CGCCTTAACAATGGCACCGCCGGGCCGAGCGGCAAATTGCTCGACCAC**TAG**CCGCTGA TTGCAGCCAAAAGACAAAACCCCGAACACCGGGGTTTTGTCTTGTCACCTCCAAGGAG CTTCCCGATGTTTGAACAGGCCAGCTGGCTCAATCAACCCCAGCATTGGCGCCGAGAA GGCGAGCGACTCAAGGTCCGCACCGATGCCAGTACCGATTTCTGGCGTGAAACCCACT ATGGTTTTGTACGCGACAACGGGCATTTCCTGTTTGTTGAAACCGACGGCGACTTTAC CGCCCAAGTCAAAATCCACAGTGAGTTTACCCACCTGTATGACCTTCGC
	ORF Start: ATG at 43 **ORF Stop: TAG at 1441**
	SEQ ID NO: 316 466 aa **MW at 49070.4kD**
NOV113a, CG59887-01 Protein Sequence	MHRTMSSFTSFALAFSMVSINTGVVTLFADPFNRVGGIGILLWLLVIPLVCCIVMVYC HLAGRIPLTGYAYQWSSRLAGNHFGWFTGWVAFTSFVAGTAATSAAIGTVFAPEIWAN PTQGQIQGLSIGATLVVGLLNICGIRLATRINDIGAIIEIIGTVLLAIALFFGVFFFF EHTQGVAILTSAQPVSGGTLSFTTIALATLLPVSVLLGWEGAADLSEETKDPRRAAPR AMIRAVLVSSVLGFVVFALLSIAIPGSVSELLSHSENPVINIVRLQLGNAAGVGMIVI AFASILACLIANMAVATRMTFALSRDNMLPGSKVLAKINPHFGTPVAAIVLITAIAVL LNLASGGFVTAIYSMVGLTYYCTYLLTLIAAYLAYKNGRMPGAPAGVFSLGRWLLPMI ILGGLWAIAVILTLSVPEESHTGAITTGVTLGVGVLWWLFSLRTRLNNGTAGPSGKLL DH
	SEQ ID NO: 317 1433 bp
NOV113b, CG59887-02 DNA Sequence	AAAAATGCACCGCACCATGAGCTCGTTCACCTCGTTTGCCCTGGCCTTTTCCATGGTC TCGATCAACACCGGCGTGGTCACGCTGTTCGCCGACCCGTTCAACCGCGTCGGGGGCA TCGGCATCCTCCTGTGGCTGTTGGTGATCCCGCTGGTGTGCTGCATCGTCATGGTCTA CTGCCACCTGGCCGGGCGCATTCCGCTCACCGGCTACGCCTACCAATGGTCCAGCCGA TTGGCGGGCAATCACTTCGGCTGGTTTACCGGCTGGGTGGCGTTCACCTCGTTTGTCG CCGGTACAGCCGCCACCTCGGCGGCCATCGGTACGGTGTTCGCACCGGAGATCTGGGC CAACCCGACACAGGGTCAGATCCAGGGCCTGAGCATCGGCGCCACGCTGGTGGTGGGC TTGCTGAATATCTGCGGGATTCGCCTGGCCACCCGGATCAACGACATCGGCGCGATCA TCGAAATCATCGGCACGGTACTGCTGGCGATTGCGTTGTTCTTCGGGGTGTTTTTCTT CTTTGAGCACACCCAGGGCGTGGCGATCCTGACCTCCGCGCAACCAGTGAGCGGCGGC ACGCTCAGCTTCACCACCATCGCCCTCGCCACCTTGCTGCCGGTCTCGGTGCTGCTGG GTTGGGAAGGCGCCGCCGACCTGTCCGAGGAAACCAAGGACCCACGCCGCGCCGCGCC CCGGGCGATGATTCGTGCGGTGCTGGTGTCCAGCGTATTGGGCTTCGTGGTGTTCGCC TTGCTGAGCATCGCGATCCCGGGCTCGGTCAGCGAACTGCTCAGCCGCAGCGAAAACC CGGTGATCAATATCGTGCGCCTGCAACTGGGCAATGCCGCCGGCGTGGGCATGGTCGT GATCGCTTTCGCCTCGATCCTCGCCTGCCTGATCGCCAACATGGCGGTGGCCACGCGC ATGACCTTCGCCCTGTCCCGGGACAACATGCTGCCGGGCTCCAAGGTGCTGGCGAAGA TCAACCCGCACTTCGGCACGCCGGTCGCCGCCATCGTGCTGATCACCGCCATCGCCGT GCTGCTGAACCTGGCGAGTGGCGGGTTTGTCACGGCGATCTACTCGATGGTCGGCCTG ACCTACTACTGCACTTACCTGCTGACGCTGATTGCCGCGTACCTGGCCTATAAAAACG GCCGGATGCCGGGGGCGCCTGCGGGCGTGTTCAGCCTGGGCCGCTGGTTGCTGCCGAT GATTATCCTCGGCGGCCTGTGGGCCATCGCGGTGATCCTGACCCTGAGCGTGCCGGAA GAAAGCCACACTGGCGCTATCACCACCGGGGTTACACTCGGCGTGGGCGTGTTGTGGT GGTTGTTTTCACTGCGCACGCGCCTTAACAATGGCACCGCCGGGCCGAGCGGCAAATT GCTCGACCAC**TAG**CCGCTGATTGCAGCCAAAAGACAAAACC
	ORF Start: ATG at 5 **ORF Stop: TAG at 1403**
	SEQ ID NO: 318 466 aa **MW at 49075.4kD**
NOV113b, CG59887-02 Protein Sequence	MHRTMSSFTSFALAFSMVSINTGVVTLFADPFNRVGGIGILLWLLVIPLVCCIVMVYC HLAGRIPLTGYAYQWSSRLAGNHFGWFTGWVAFTSFVAGTAATSAAIGTVFAPEIWAN PTQGQIQGLSIGATLVVGLLNICGIRLATRINDIGAIIEIIGTVLLAIALFFGVFFFF EHTQGVAILTSAQPVSGGTLSFTTIALATLLPVSVLLGWEGAADLSEETKDPRRAAPR AMIRAVLVSSVLGFVVFALLSIAIPGSVSELLSRSENPVINIVRLQLGNAAGVGMVVI AFASILACLIANMAVATRMTFALSRDNMLPGSKVLAKINPHFGTPVAAIVLITAIAVL LNLASGGFVTAIYSMVGLTYYCTYLLTLIAAYLAYKNGRMPGAPAGVFSLGRWLLPMI ILGGLWAIAVILTLSVPEESHTGAITTGVTLGVGVLWWLFSLRTRLNNGTAGPSGKLL DH

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 113B.

Table 113B. Comparison of NOV113a against NOV113b.		
Protein Sequence	NOV113a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV113b	1..466 1..466	343/466 (73%) 344/466 (73%)

Further analysis of the NOV113a protein yielded the following properties shown in Table 113C.

Table 113C. Protein Sequence Properties NOV113a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 59 and 60

A search of the NOV113a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 113D.

Table 113D. Geneseq Results for NOV113a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV113a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG49885	Arabidopsis thaliana protein fragment SEQ ID NO: 63155 - Arabidopsis thaliana, 504 aa. [EP1033405-A2, 06-SEP-2000]	1..449 17..492	122/486 (25%) 217/486 (44%)	3e-31
AAG49884	Arabidopsis thaliana protein fragment SEQ ID NO: 63154 - Arabidopsis thaliana, 516 aa. [EP1033405-A2, 06-SEP-2000]	1..449 29..504	122/486 (25%) 217/486 (44%)	3e-31
AAG20282	Arabidopsis thaliana protein fragment SEQ ID NO: 22407 - Arabidopsis thaliana, 504 aa. [EP1033405-A2, 06-SEP-2000]	1..449 17..492	122/486 (25%) 217/486 (44%)	3e-31
AAG20281	Arabidopsis thaliana protein fragment SEQ ID NO: 22406 - Arabidopsis thaliana, 516 aa. [EP1033405-A2, 06-SEP-2000]	1..449 29..504	122/486 (25%) 217/486 (44%)	3e-31

AAG20280	Arabidopsis thaliana protein fragment SEQ ID NO: 22405 - Arabidopsis thaliana, 528 aa. [EP1033405-A2, 06-SEP-2000]	1..449 41..516	122/486 (25%) 217/486 (44%)	3e-31

In a BLAST search of public sequence databases, the NOV113a protein was found to have homology to the proteins shown in the BLASTP data in Table 113E.

Table 113E. Public BLASTP Results for NOV113a				
Protein Accession Number	Protein/Organism/Length	NOV113a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9KZF1	PROBABLE AMINO ACID/METABOLITE PERMEASE - Streptomyces coelicolor, 504 aa.	3..450 27..481	139/469 (29%) 214/469 (44%)	2e-41
Q98H14	AMINO ACID/METABOLITE PERMEASE - Rhizobium loti (Mesorhizobium loti), 518 aa.	1..446 27..485	118/466 (25%) 209/466 (44%)	1e-36
Q92NI8	PUTATIVE AMINO-ACID PERMEASE PROTEIN - Rhizobium meliloti (Sinorhizobium meliloti), 515 aa.	1..449 25..487	122/475 (25%) 204/475 (42%)	1e-32
O22509	PUTATIVE AMINO ACID OR GABA PERMEASE - Arabidopsis thaliana (Mouse-ear cress), 516 aa.	1..449 29..504	122/486 (25%) 217/486 (44%)	1e-30
Q9ZU50	PUTATIVE AMINO ACID PERMEASE - Arabidopsis thaliana (Mouse-ear cress), 517 aa.	1..449 29..505	120/487 (24%) 216/487 (43%)	2e-28

PFam analysis predicts that the NOV113a protein contains the domains shown in the Table 113F.

5

Table 113F. Domain Analysis of NOV113a			
Pfam Domain	NOV113a Match Region	Identities/ Similarities for the Matched Region	Expect Value
oxidored_q3: domain 1 of 1	162..307	28/182 (15%) 91/182 (50%)	3.7

ISK_Channel: domain 1 of 1	196..326	32/136 (24%) 55/136 (40%)	8.8
ABC2_membrane: domain 1 of 1	122..377	46/273 (17%) 154/273 (56%)	8.3
SSF: domain 1 of 1	7..394	77/470 (16%) 222/470 (47%)	7.8
Aa_trans: domain 1 of 1	29..417	67/483 (14%) 236/483 (49%)	9.7
aa_permeases: domain 1 of 1	1..451	86/516 (17%) 287/516 (56%)	1.1e-05

Example 114.

The NOV114 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 114A.

Table 114A. NOV114 Sequence Analysis			
	SEQ ID NO: 319	876 bp	
NOV114a, CG59861-01 DNA Sequence	AACTTGCTTTTGGGAGCCAGCGGTATGGCGTCGGGCTGCAAGATTGGCCCGTCCATCC TCAACAGCGACCTGGCCAATTTAGGGGCCGAGTGCTCCCGGATGCTAGACTCTGGGGC CGATTATCTGCACCTGGACGTAATGGACGGGCATTTTGTTCCCAACATCACCTTTGGT CACCCTGTGGTGGAAAGCCTTCGAAAGCAGCTAGGCCAGGACCCTTTCTTTGACATGC ACATGATGGTGTCCAAGCCAGAACAGTGGGTAAAGCCAATGGCTGTAGCAGGAGCCAA TCAGTACACCTTTCATCTCGAGGCTACTGAGAACCCAGGGGCTTTGATTAAAGACATT CGGGAGAATGGGATGAAGGTTGGCCTTGCCATCAAACCAGGAACCTCAGTTGAGTATT TGGCACCATGGGCTAATCAGATAGATATGGCCTTGGTTATGACAGTGGAACCGGGGTT TGGAGGGCAGAAATTCATGGAAGATATGATGCCAAAGGTTCACTGGTTGAGGACCCAG TTCCCATCTTTGGATATAGAGGTCGATGGTGGAGTAGGTCCTGACACTGTCCATAAAT GTGCAGAGGCAGGAGCTAACATGATTGTGTCTGGCAGTGCTATTATGAGGAGTGAAGA CCCCAGATCTGTGATCAATCTATTAAGAAATGTTTGCTCAGAAGCTGCTCAGAAACGT TCTCTTGATCGGTGAAACCATAAGGAGCCCAGTGTTCCTGTTCATGAAATCTCCCTTT TACTGGAAAACAGGAATATTGACTACCAAATCACAATGCAATTGAAGCCGTACTGCTT TTTTGAGCAGTTATTCATTCCAGTGATTAAAACTGATTGTGCAGAATAAAAAAAAAAA AAAAAA
	ORF Start: ATG at 25	ORF Stop: TGA at 709	
	SEQ ID NO: 320	228 aa	MW at 24901.4kD
NOV114a, CG59861-01 Protein Sequence	MASGCKIGPSILNSDLANLGAECSRMLDSGADYLHLDVMDGHFVPNITFGHPVVESLR KQLGQDPFFDMHMMVSKPEQWVKPMAVAGANQYTFHLEATENPGALIKDIRENGMKVG LAIKPGTSVEYLAPWANQIDMALVMTVEPGFGGQKFMEDMMPKVHWLRTQFPSLDIEV DGGVGPDTVHKCAEAGANMIVSGSAIMRSEDPRSVINLLRNVCSEAAQKRSLDR		
	SEQ ID NO: 321	730 bp	
NOV114b, CG59861-02 DNA Sequence	AACTTGCTTTTGGGAGCCAGCGGTATGGCGTCGGGCTGCAAGATTGGCCCGTCCATCC TCAACAGCGACCTGGCCAATTTAGGGGCCGAGTGCTCCCGGATGCTAGACTCTGGGGC CGATTATCTGCACCTGGACGTAATGGACGGGCATTTTGTTCCCAACATCACCTTTGGT CACCCTGTGGTAGAAAGCCTTCGAAAGCAGCTAGGCCAGGACCCTTTCTTTGACATGC ACATGATGGTGTCCAAGCCAGAACAGTGGGTAAAGCCAATGGCTGTAGCAGGAGCCAA TCAGTACACCTTTCATCTCGAGGCTACTGAGAACCCAGGGGCTTTGATTAAAGACATT CGGGAGAATGGGATGAAGGTTGGCCTTGCCATCAAACCAGGAACCTCAGTTGAGTATT TGGCACCATGGGCTAATCAGATAGATATGGCCTTGGTTATGACAGTGGAACCGGGGTT TGGAGGGCAGAAATTCATGGAAGATATGATGCCAAAGGTTCACTGGTTGAGGACCCAG TTCCCATCTTTGGATATAGAGGTCGATGGTGGAGTAGGTCCTGACACTGTCCATAAAT GTGCAGAGGCAGGAGCTAACATGATTGTGTCTGGCAGTGCTATTATGAGGAGTGAAGA CCCCAGATCTGTGATCAATCTATTAAGAAATGTTTGCTCAGAAGCTGCTCAGAAACGT TCTCTTGATCGGTGAAACCATAAGGAGCCCAGTG
	ORF Start: ATG at 25	ORF Stop: TGA at 709	
	SEQ ID NO: 322	228 aa	MW at 24927.5kD

NOV114b, CG59861-02 Protein Sequence	MASGCKIGPSILNSDLANLGAECLRMLDSGADYLHLDVMDGHFVPNITFGHPVVESLR KQLGQDPFFDMHMMVSKPEQWVKPMAVAGANQYTFHLEATENPGALIKDIRENGMKVG LAIKPGTSVEYLAPWANQIDMALVMTVEPGFGGQKFMEDMMPKVHWLRTQFPSLDIEV DGGVGPDTVHKCAEAGANMIVSGSAIMRSEDPRSVINLLRNVCSEAAQKRSLDR

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 114B.

Table 114B. Comparison of NOV114a against NOV114b.

Protein Sequence	NOV114a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV114b	1..228 1..228	227/228 (99%) 227/228 (99%)

Further analysis of the NOV114a protein yielded the following properties shown in Table 114C.

Table 114C. Protein Sequence Properties NOV114a

PSort analysis:	0.6500 probability located in cytoplasm; 0.1753 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space; 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV114a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 114D.

Table 114D. Geneseq Results for NOV114a

Geneseq Identifier	Protein/Organism/Length [Patent #, Protein/Organism/Length [Patent #, Date]	NOV114a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41358	Human polypeptide SEQ ID NO 6289 - Homo sapiens, 247 aa. [WO200153312-A1, 26-JUL-2001]	1..228 20..247	227/228 (99%) 227/228 (99%)	e-132
AAM41357	Human polypeptide SEQ ID NO 6288 - Homo sapiens, 247 aa. [WO200153312-A1, 26-JUL-2001]	1..228 20..247	227/228 (99%) 227/228 (99%)	e-132

AAM39571	Human polypeptide SEQ ID NO 2716 - Homo sapiens, 228 aa. [WO200153312-A1, 26-JUL-2001]	1..228 1..228	227/228 (99%) 227/228 (99%)	e-132
AAB71912	Human ISOM-4 - Homo sapiens, 228 aa. [WO200112790-A2, 22-FEB-2001]	1..228 1..228	227/228 (99%) 227/228 (99%)	e-132
AAM39572	Human polypeptide SEQ ID NO 2717 - Homo sapiens, 246 aa. [WO200153312-A1, 26-JUL-2001]	1..228' 1..246	227/246 (92%) 227/246 (92%)	e-129

In a BLAST search of public sequence databases, the NOV114a protein was found to have homology to the proteins shown in the BLASTP data in Table 114E.

Table 114E. Public BLASTP Results for NOV114a				
Protein Accession Number	Protein/Organism/Length	NOV114a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96AT9	HYPOTHETICAL 24.9 KDA PROTEIN - Homo sapiens (Human), 228 aa.	1..228 1..228	227/228 (99%) 227/228 (99%)	e-131
Q9BSB5	HYPOTHETICAL 25.3 KDA PROTEIN - Homo sapiens (Human), 232 aa (fragment).	1..228 5..232	227/228 (99%) 227/228 (99%)	e-131
AAH19126	HYPOTHETICAL 24.9 KDA PROTEIN - Mus musculus (Mouse), 228 aa.	1..228 1..228	221/228 (96%) 226/228 (98%)	e-129
O43767	RIBULOSE-5-PHOSPHATE-EPIMERASE - Homo sapiens (Human), 174 aa (fragment).	55..228 1..174	174/174 (100%) 174/174 (100%)	2e-98
Q96N34	CDNA FLJ31466 FIS, CLONE NT2NE2001372, HIGHLY SIMILAR TO HOMO SAPIENS PUTATIVE RIBULOSE-5-PHOSPHATE-EPIMERASE - Homo sapiens (Human), 178 aa.	69..228 1..178	160/178 (89%) 160/178 (89%)	2e-86

PFam analysis predicts that the NOV114a protein contains the domains shown in
5   the Table 114F.

Table 114F. Domain Analysis of NOV114a			
Pfam Domain	**NOV114a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Ribul_P_3_epim: domain 1 of 1	6..204	95/209 (45%) 174/209 (83%)	1.9e-105
IGPS: domain 1 of 1	179..213	15/35 (43%) 27/35 (77%)	0.02
trp_syntA: domain 1 of 1	34..222	45/273 (16%) 124/273 (45%)	2.9

Example 115.

The NOV115 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 115A.

Table 115A. NOV115 Sequence Analysis			
	SEQ ID NO: 323	1761 bp	
NOV115a, CG59857-01 DNA Sequence	AGTGTGGTACCTATCTGTCCCCCCTCTGGAGGGGTTGACAAGGGAAAGGGCACCGGGG GGCACAGAGA**TG**CAGGACAGATTGCACATCCTGGAGGACCTGAATATGCTCTACATTC GGCAGATGGCACTCAGCCTGGAGGACACGGAGTTGCAGAGGAAGCTAGACCATGAGAT CCGGATGAGGGAAGGGGCCTGTAAGCTGCTGGCAGCCTGCTCCCAGCGAGAGCAGGCT CTGGAGGCCACCAAGAGCCTGCTAGTGTGCAACAGCCGCATCCTCAGCTACATGGGCG AGCTGCAGCGGCGCAAGGAGGCGCAGGTGCTGGGGAAGACAAGCCGGCGGCCTTCTGA CAGTGGCCCGCCCGCTGAGCGCTCCCCCTGCCGCGGCCGGGTCTGCATCTCTGACCTC CGGATTCCACTCATGTGGAAGGACACAGAATATTTCAAGAACAAAGACTTGCACCGCT GGGCTGTGTTCCTGCTGCTGCAGCTGGGGGAACACATCCAGGACACAGAGATGATCCT AGTGGACAGGACCCTCACAGACATCTCCTTTCAGAGCAATGTGCTCTTCGCTGAGGCG GGGCCAGACTTTGAACTGCGGTTAGAGCTGTATGGGGCCTGTGTGGAAGAAGAGGGGG CCCTGACTGGCGGCCCCAAGAGGCTTGCCACCAAACTCAGCAGCTCCCTGGGCCGCTC CTCAGGGAGGCGTGTCCGGGCATCGCTGGACAGTGCTGGGGGTTCAGGGAGCAGTCCC ATCTTGCTCCCCACCCCAGTTGTTGGTGGTCCTCGTTACCACCTCTTGGCTCACACCA CACTCACCCTGGCAGCAGTGCAAGATGGATTCCGCACACATGACCTCACCCTTGCCAG TCATGAGGAGAACCCTGCCTGGCTGCCCCTTTATGGTAGCGTGTGTTGCCGTCTGGCA GCTCAGCCTCTCTGCATGACTCAGCCCACTGCAAGTGGTACCCTCAGGGTGCAGCAAG CTGGGGAGATGCAGAACTGGGCACAAGTGCATGGAGTTCTGAAAGGCACAAACCTCTT CTGTTACCGGCAACCTGAGGATGCAGACACTGGGGAAGAGCCGCTGCTTACTATTGCT GTCAACAAGGAGACTCGAGTCCGGGCAGGGGAGCTGGACCAGGCTCTAGGACGGCCCT TCACCCTAAGCATCAGTAACCAGTATGGGGATGATGAGGTGACACACACCCTTCAGAC AGAAAGTCGGGAAGCACTGCAGAGCTGGATGGAGGCTCTGTGGCAGCTTTTCTTTGAC ATGAGCCAATGGAAGCAGTGCTGTGATGAAATCATGAAAATTGAAACTCCTGCTCCCC GGAAACCACCCCAAGCACTGGCAAAGCAGGGGTCCTTGTACCATGAGATGGCTATTGA GCCGCTGGATGACATCGCAGCGGTGACAGACATCCTGACCCAGCGGGAGGGCGCAAGG CTGGAGACACCCCCACCCTGGCTGGCAATGTTTACAGACCAGCCTGCCCTGCCTAACC CCTGCTCGCCTGCCTCAGTGGCCCCAGCCCCAGACTGGACCCACCCCCTGCCCTGGGG GAGACCCCGAACCTTTTCCCTGGATGCTGTCCCCCCAGACCACTCCCCTAGGGCTCGC TCGGTTGCCCCCCTCCCACCTCAGCGATCCCCACGGACCAGAGGCCTCTGCAGCAAAG GCCAACCTCGCACTTGGCTCCAGTCACCAGTG**TGA**GAGAGAAAGGTGCTGGCATAGGA TCTGCCCAGAAGAGAAAATGA
	ORF Start: ATG at 68	ORF Stop: TGA at 1715	
	SEQ ID NO: 324	549 aa	MW at 61171.0kD
NOV115a, CG59857-01 Protein Sequence	MQDRLHILEDLNMLYIRQMALSLEDTELQRKLDHEIRMREGACKLLAACSQREQALEA TKSLLVCNSRILSYMGELQRRKEAQVLGKTSRRPSDSGPPAERSPCRGRVCISDLRIP LMWKDTEYFKNKDLHRWAVFLLLQLGEHIQDTEMILVDRTLTDISFQSNVLFAEAGPD FELRLELYGACVEEEGALTGGPKRLATKLSSSLGRSSGRRVRASLDSAGGSGSSPILL PTPVVGGPRYHLLAHTTLTLAAVQDGFRTHDLTLASHEENPAWLPLYGSVCCRLAAQP LCMTQPTASGTLRVQQAGEMQNWAQVHGVLKGTNLFCYRQPEDADTGEEPLLTIAVNK ETRVRAGELDQALGRPFTLSISNQYGDDEVTHTLQTESREALQSWMEALWQLFFDMSQ

441

WKQCCDEIMKIETPAPRKPPQALAKQGSLYHEMAIEPLDDIAAVTDILTQREGARLET PPPWLAMFTDQPALPNPCSPASVAPAPDWTHPLPWGRPRTFSLDAVPPDHSPRARSVA PLPPQRSPRTRGLCSKGQPRTWLQSPV

Further analysis of the NOV115a protein yielded the following properties shown in Table 115B.

Table 115B. Protein Sequence Properties NOV115a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1707 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV115a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 115C.

Table 115C. Geneseq Results for NOV115a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV115a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB35241	Human rhotekin - Homo sapiens, 563 aa. [US6183990-B1, 06-FEB-2001]	24..549 37..563	526/527 (99%) 526/527 (99%)	0.0
AAY44559	Human Rhotekin protein - Homo sapiens, 563 aa. [WO9958667-A1, 18-NOV-1999]	24..549 37..563	526/527 (99%) 526/527 (99%)	0.0
AAB35242	Human rhotekin EST-derived protein - Homo sapiens, 527 aa. [US6183990-B1, 06-FEB-2001]	24..549 1..527	522/527 (99%) 523/527 (99%)	0.0
AAY44560	Human Rhotekin variant protein - Homo sapiens, 527 aa. [WO9958667-A1, 18-NOV-1999]	24..549 1..527	522/527 (99%) 523/527 (99%)	0.0
AAB26790	Human Ras correlative GTP binding kinase protein sequence - Homo sapiens, 544 aa. [CN1257924-A, 28-JUN-2000]	24..549 18..544	518/527 (98%) 519/527 (98%)	0.0

In a BLAST search of public sequence databases, the NOV115a protein was found to have homology to the proteins shown in the BLASTP data in Table 115D.

Table 115D. Public BLASTP Results for NOV115a				
Protein Accession Number	Protein/Organism/Length	NOV115a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH17727	SIMILAR TO RHOTEKIN - Homo sapiens (Human), 550 aa.	1..549 1..550	549/550 (99%) 549/550 (99%)	0.0
Q9BST9	SIMILAR TO RHOTEKIN - Homo sapiens (Human), 587 aa (fragment).	24..549 61..587	526/527 (99%) 526/527 (99%)	0.0
Q96PT6	RTKN - Homo sapiens (Human), 544 aa.	24..549 18..544	518/527 (98%) 519/527 (98%)	0.0
Q9HB05	RHOTEKIN - Homo sapiens (Human), 567 aa (fragment).	24..549 41..567	505/527 (95%) 513/527 (96%)	0.0
Q61192	RHOTEKIN - Mus musculus (Mouse), 551 aa.	1..549 1..551	477/551 (86%) 500/551 (90%)	0.0

PFam analysis predicts that the NOV115a protein contains the domains shown in the Table 115E.

Table 115E. Domain Analysis of NOV115a			
Pfam Domain	NOV115a Match Region	Identities/ Similarities for the Matched Region	Expect Value
HR1: domain 1 of 1	23..95	17/87 (20%) 54/87 (62%)	0.27
PH: domain 1 of 1	296..397	19/102 (19%) 72/102 (71%)	1e-06

Example 116.

The NOV116 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 116A.

Table 116A. NOV116 Sequence Analysis		
	SEQ ID NO: 325	450 bp
NOV116a, CG59855-01 DNA Sequence	CTGGGAGACTGAAAAAATGCAGACCACCGGGGTATTACTCATTTCTCCAGCTCTGATC TGCTGTTGTACCAGGGGTCTAATCAGGCCTGTGTCTGCCTTCTCCTTGAATAGCCCAG AGAATTCATCTAAACAGCCTTCCTACAGCAGCTCCCCACTCCAGGTGGCCAGACGGGA GTTCCAGACCAGTGTTGTCTCCCGGGACACTGACACAGCCGCCAAGTTTATTGGTGCT GGGTCAGCCACAGTTGGTGTGGCTGATTCAGGGGCTGGCATTGGAGCGGTGTTTGGCA GCTTGATTATTGTCTATGCCAGGAAGCTGTCTCTCAAGCAGCAACTCCTCTTCTATGC CATTCTGGGCTTTGCCCTGTCTGAGGCCATGGGGCTCTTCTGTTTGATGATCTCCTTC TTCATCCTGTTCGCCATGTGAGGCTCCGTGAGGGTCACCTGCCT	

443

	ORF Start: ATG at 17	ORF Stop: TGA at 425	
	SEQ ID NO: 326	136 aa	MW at 14384.6kD
NOV116a, CG59855-01 Protein Sequence	MQTTGVLLISPALICCCTRGLIRPVSAFSLNSPENSSKQPSYSSSPLQVARREFQTSV VSRDTDTAAKFIGAGSATVGVADSGAGIGAVFGSLIIVYARKLSLKQQLLFYAILGFA LSEAMGLFCLMISFFILFAM		
	SEQ ID NO: 327	434 bp	
NOV116b, CG59855-02 DNA Sequence	ATGCAGACCACCGGGGTATTACTCATTTCTCCAGCTCTGATCTGCTGTTGTACCAGGG GTCTAATCAGGCCTGTGTCTGCCTTCTCCTTGAATAGCCCAGAGAATTCATCTAAACA GCCTTCCTACAGCAGCTCCCCACTCCAGGTGGCCAGACGGGAGTTCCAGACCAGTGTT GTCTCCCGGGACACTGACACAGCCGCCAAGTTTATTGGTGCTGGGTCAGCCACAGTTG GTGTGGCTGATTCAGAGGCTGGCATTGGAGCGGTGTTTGGCAGCTTGATTATTGTCTA TGCCAGGAAGCTGTCTCTCAAGCAGCAACTCCTCTTCTATGCCATTCTGGGCTTTGCC CTGTCTGAGGCCATGGGGCTCTTCTGTTTGATGATCTCCTTCTTCATCCTGTTCGCCA TGTGAGGCTCCGTGAGGGTCACCTGCCT		
	ORF Start: ATG at 1	ORF Stop: TGA at 409	
	SEQ ID NO: 328	136 aa	MW at 14456.7kD
NOV116b, CG59855-02 Protein Sequence	MQTTGVLLISPALICCCTRGLIRPVSAFSLNSPENSSKQPSYSSSPLQVARREFQTSV VSRDTDTAAKFIGAGSATVGVADSEAGIGAVFGSLIIVYARKLSLKQQLLFYAILGFA LSEAMGLFCLMISFFILFAM		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 116B.

Table 116B. Comparison of NOV116a against NOV116b.		
Protein Sequence	NOV116a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV116b	1..136 1..136	120/136 (88%) 120/136 (88%)

Further analysis of the NOV116a protein yielded the following properties shown in Table 116C.

Table 116C. Protein Sequence Properties NOV116a	
PSort analysis:	0.9190 probability located in plasma membrane; 0.3000 probability located in lysosome (membrane); 0.1888 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Likely cleavage site between residues 28 and 29

A search of the NOV116a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 116D.

444

Table 116D. Geneseq Results for NOV116a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV116a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG75142	Human colon cancer antigen protein SEQ ID NO:5906 - Homo sapiens, 142 aa. [WO200122920-A2, 05-APR-2001]	1..136 7..142	115/136 (84%) 119/136 (86%)	2e-57
AAB43866	Human cancer associated protein sequence SEQ ID NO:1311 - Homo sapiens, 142 aa. [WO200055350-A1, 21-SEP-2000]	1..136 7..142	115/136 (84%) 119/136 (86%)	2e-57
AAU69713	Cell death protective sequence CNI-00730, protein #1 - Homo sapiens, 142 aa. [WO200176532-A2, 18-OCT-2001]	7..136 7..142	85/136 (62%) 98/136 (71%)	2e-36
ABB12016	Human ATP synthase subunit homologue, SEQ ID NO:2386 - Homo sapiens, 187 aa. [WO200157188-A2, 09-AUG-2001]	7..136 52..187	85/136 (62%) 98/136 (71%)	2e-36
AAB53428	Human colon cancer antigen protein sequence SEQ ID NO:968 - Homo sapiens, 212 aa. [WO200055351-A1, 21-SEP-2000]	7..136 77..212	85/136 (62%) 98/136 (71%)	2e-36

In a BLAST search of public sequence databases, the NOV116a protein was found to have homology to the proteins shown in the BLASTP data in Table 116E.

Table 116E. Public BLASTP Results for NOV116a				
Protein Accession Number	Protein/Organism/Length	NOV116a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P05496	ATP synthase lipid-binding protein, mitochondrial precursor (EC 3.6.1.34) (ATP synthase proteolipid P1) (ATPase protein 9) (ATPase subunit C) - Homo sapiens (Human), 136 aa.	1..136 1..136	115/136 (84%) 119/136 (86%)	9e-57

P32876	ATP synthase lipid-binding protein, mitochondrial precursor (EC 3.6.1.34) (ATP synthase proteolipid P1) (ATPase protein 9) (ATPase subunit C) - Bos taurus (Bovine), 136 aa.	1..136 1..136	113/136 (83%) 117/136 (85%)	1e-54
P17605	ATP synthase lipid-binding protein, mitochondrial precursor (EC 3.6.1.34) (ATP synthase proteolipid P1) (ATPase protein 9) (ATPase subunit C) - Ovis aries (Sheep), 136 aa.	1..136 1..136	113/136 (83%) 117/136 (85%)	2e-54
Q9CR84	ATP SYNTHASE C CHAIN ISOFORM 1 (EC 3.6.1.34) (LIPID-BINDING PROTEIN) (SUBUNIT C) - Mus musculus (Mouse), 136 aa.	1..136 1..136	112/136 (82%) 117/136 (85%)	1e-53
P48202	ATP synthase lipid-binding protein, mitochondrial precursor (EC 3.6.1.34) (ATP synthase proteolipid P1) (ATPase protein 9) (ATPase subunit C) - Mus musculus (Mouse), 136 aa.	1..136 1..136	112/136 (82%) 117/136 (85%)	1e-53

PFam analysis predicts that the NOV116a protein contains the domains shown in the Table 116F.

Table 116F. Domain Analysis of NOV116a			
Pfam Domain	NOV116a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ATP-synt_C: domain 1 of 1	67..135	31/70 (44%) 57/70 (81%)	2.3e-18

Example 117.

The NOV117 clone was analyzed, and the nucleotide and predicted polypeptide
5   sequences are shown in Table 117A.

Table 117A. NOV117 Sequence Analysis		
	SEQ ID NO: 329	1769 bp
NOV117a, CG59807-01 DNA Sequence	GAGGTGATGCTGGAGACCTGCGGACTTCTCATGTCTCTGGGCTGTCCTTTGTTCAAAC CAGAGCTGATCTACCAGTTGGATCACAGACAGGAGCTATGGATGGCTACAAAAGACCT CTCCCAAAGCTCCTATCCAGGTGACAACACAAAACCCAAGACCACAGAGCCTACCTTT TCTCACCTGGCCTTGCCTGAGGAAGTCTTACTCCAGGAACAACTGACACAAGGAGCCT CAAAGAACTCCCAATTAGGGCAATCCAAGGATCAGGATGGGCCATCTGAAATGCAAGA AGTCCACTTGAAAATAGGGATAGGCCCCCAGCGGGGGAAGCTGCTGGAGAAAATGAGT	

446

	TCTGAACGTGATGGTTTGGGGTCAGATGATGGTGTATGTACAAAGATTACACAGAAAC
	AAGTTTCAACAGAAGGTGATCTCTATGAATGTGATTCACATGGACCAGTTACAGATGC
	CTTGATTCGCGAAGAGAAAAATTCCTATAAATGTGAGGAATGCGGGAAAGTGTTTAAA
	AAGAATGCCCTCCTTGTTCAGCATGAACGGATTCACACTCAAGTGAAGCCCTATGAAT
	GCACAGAGTGTGGGAAAACCTTTAGCAAGAGCACTCATCTTCTTCAGCACCTCATCAT
	CCACACTGGGGAGAAGCCCTATAAGTGCATGGAGTGTGGGAAGGCTTTTAACCGCAGG
	TCACACCTCACACGGCACCAGCGGATTCACAGTGGAGAGAAGCCTTATAAGTGCAGTG
	AATGTGGAAAGGCCTTCACCCACCGCTCCACTTTTGTCTTGCATCACAGGAGCCACAC
	TGGAGAAAAACCCTTTGTGTGCAAAGAGTGTGGCAAAGCCTTTCGAGATAGGCCAGGT
	TTCATTCGACACTACATCATCCACACGGGAGAGAAGCCCTATGAGTGCATTGAGTGTG
	GGAAGGCCTTCAACCGCCGGTCATACCTCACGTGGCACCAACAGATTCACACTGGAGT
	GAAACCCTTTGAATGCAACGAGTGTGGAAAAGCTTTTTGCGAGAGTGCAGACCTCATT
	CAACACTACATTATCCACACTGGGGAGAAGCCCTATAAGTGCATGGAGTGTGGGAAGG
	CGTTCAACCGTAGGTCACACCTCAAGCAGCATCAACGGATTCACACTGGGGAGAAGCC
	TTATGAATGCAGTGAATGTGGAAAGGCCTTCACCCACTGCTCCACTTTTGTCTTGCAT
	AAAAGGACCCACACAGGAGAAAAACCCTATGAATGCAAAGAATGTGGAAAAGCCTTTA
	GTGATAGGGCAGACCTCATTCGCCACTTCAGCATCCACACTGGAGAGAAACCCTATGA
	GTGCGTGGAGTGTGGAAAGGCCTTCAACCGCAGCTCACACCTCACGAGGCACCAACAG
	ATTCACACTGGAGAGAAACCCTATGAATGCATCCAGTGTGGGAAAGCCTTTTGCCGGA
	GCGCAAACCTTATTCGACACTCCATCATTCACACTGGAGAGAAGCCGTATGAATGCAG
	TGAGTGTGGAAAGGCTTTTAATCGCGGCTCATCCCTCACACATCATCAAAGGATTCAT
	ACTGGGAGAAACCCTACCATTGTAACAGATGTGGGAAGACCTTTTATGACTGCACAGA
	CTTCAGTCAACATCCAGGAACTTTTATTAGGGAAAGAGTTTTTGAATATCACCACTGA
	AGAAAATCGTGGTGAAAGGGAACATCTTACCATCTGGCCATTCACACTGAAGAGAAA
	CTTCATAAGCATCCTCTCTTTGAGAAAAC
ORF Start: ATG at 7	**ORF Stop: TGA at 1696**
SEQ ID NO: 330	**563 aa** **MW at 64300.6kD**
NOV117a, CG59807-01 Protein Sequence	MLETCGLLMSLGCPLFKPELIYQLDHRQELWMATKDLSQSSYPGDNTKPKTTEPTFSH LALPEEVLLQEQLTQGASKNSQLGQSKDQDGPSEMQEVHLKIGIGPQRGKLLEKMSSE RDGLGSDDGVCTKITQKQVSTEGDLYECDSHGPVTDALIREEKNSYKCEECGKVFKKN ALLVQHERIHTQVKPYECTECGKTFSKSTHLLQHLIIHTGEKPYKCMECGKAFNRRSH LTRHQRIHSGEKPYKCSECGKAFTHRSTFVLHHRSHTGEKPFVCKECGKAFRDRPGFI RHYIIHTGEKPYECIECGKAFNRRSYLTWHQQIHTGVKPFECNECGKAFCESADLIQH YIIHTGEKPYKCMECGKAFNRRSHLKQHQRIHTGEKPYECSECGKAFTHCSTFVLHKR THTGEKPYECKECGKAFSDRADLIRHFSIHTGEKPYECVECGKAFNRSSHLTRHQQIH TGEKPYECIQCGKAFCRSANLIRHSIIHTGEKPYECSECGKAFNRGSSLTHHQRIHTG RNPTIVTDVGRPFMTAQTSVNIQELLLGKEFLNITTEENLW

Further analysis of the NOV117a protein yielded the following properties shown in Table 117B.

Table 117B. Protein Sequence Properties NOV117a	
Psort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Likely cleavage site between residues 19 and 20

A search of the NOV117a protein against the Geneseq database, a proprietary
database that contains sequences published in patents and patent publication, yielded
several homologous proteins shown in Table 117C.

Table 117C. Geneseq Results for NOV117a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV117a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM79549	Human protein SEQ ID NO 3195 - Homo sapiens, 603 aa. [WO200157190-A2, 09-AUG-2001]	1..563 38..603	563/566 (99%) 563/566 (99%)	0.0
AAM78565	Human protein SEQ ID NO 1227 - Homo sapiens, 603 aa. [WO200157190-A2, 09-AUG-2001]	1..563 38..603	563/566 (99%) 563/566 (99%)	0.0
ABB21767	Protein #3766 encoded by probe for measuring heart cell gene expression - Homo sapiens, 551 aa. [WO200157274-A2, 09-AUG-2001]	44..562 10..527	375/519 (72%) 437/519 (83%)	0.0
AAM69575	Human bone marrow expressed probe encoded protein SEQ ID NO: 29881 - Homo sapiens, 551 aa. [WO200157276-A2, 09-AUG-2001]	44..562 10..527	375/519 (72%) 437/519 (83%)	0.0
AAM57172	Human brain expressed single exon probe encoded protein SEQ ID NO: 29277 - Homo sapiens, 551 aa. [WO200157275-A2, 09-AUG-2001]	44..562 10..527	375/519 (72%) 437/519 (83%)	0.0

In a BLAST search of public sequence databases, the NOV117a protein was found to have homology to the proteins shown in the BLASTP data in Table 117D.

Table 117D. Public BLASTP Results for NOV117a				
Protein Accession Number	Protein/Organism/Length	NOV117a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O43296	Zinc finger protein 264 - Homo sapiens (Human), 627 aa.	1..562 43..603	401/562 (71%) 468/562 (82%)	0.0
Q96NL3	CDNA FLJ30663 FIS, CLONE	1..535 38..572	299/535 (55%) 369/535 (68%)	0.0

448

	SIMILAR TO ZINC FINGER PROTEIN 84 - Homo sapiens (Human), 588 aa.			
Q99676	Zinc finger protein 184 - Homo sapiens (Human), 751 aa.	2..535 58..595	261/542 (48%) 355/542 (65%)	e-151
Q96SE7	ZINC FINGER 1111 - Homo sapiens (Human), 839 aa.	151..541 306..694	233/391 (59%) 281/391 (71%)	e-148
Q03923	Zinc finger protein 85 (Zinc finger protein HPF4) (HTF1) - Homo sapiens (Human), 595 aa.	1..535 33..547	266/544 (48%) 328/544 (59%)	e-148

PFam analysis predicts that the NOV117a protein contains the domains shown in the Table 117E.

Table 117E. Domain Analysis of NOV117a			
Pfam Domain	NOV117a Match Region	Identities/ Similarities for the Matched Region	Expect Value
KRAB: domain 1 of 1	1..34	14/66 (21%) 24/66 (36%)	0.15
zf-C2H2: domain 1 of 13	162..184	11/24 (46%) 19/24 (79%)	3.6e-06
zf-C2H2: domain 2 of 13	190..212	11/24 (46%) 19/24 (79%)	7.1e-06
zf-C2H2: domain 3 of 13	218..240	14/24 (58%) 22/24 (92%)	2.3e-07
zf-BED: domain 1 of 3	203..241	13/52 (25%) 25/52 (48%)	2
zf-C2H2: domain 4 of 13	246..268	11/24 (46%) 20/24 (83%)	4.6e-05
LIM: domain 1 of 1	220..284	16/72 (22%) 50/72 (69%)	0.69
zf-C2H2: domain 5 of 13	274..296	8/24 (33%) 18/24 (75%)	7.6e-05
zf-C2H2: domain 6 of 13	302..324	11/24 (46%) 20/24 (83%)	8.4e-05
Zn_carbOpept: domain 1 of 1	312..330	5/19 (26%) 17/19 (89%)	1.2

449

zf-C2H2: domain 7 of 13	330..352	8/24 (33%) 19/24 (79%)	9.7e-05
zf-C2H2: domain 8 of 13	358..380	14/24 (58%) 22/24 (92%)	5.3e-07
zf-BED: domain 2 of 3	343..381	12/52 (23%) 26/52 (50%)	1.3
zf-C2H2: domain 9 of 13	386..408	11/24 (46%) 20/24 (83%)	9.4e-05
zf-C2H2: domain 10 of 13	414..436	11/24 (46%) 20/24 (83%)	5e-06
zf-C2H2: domain 11 of 13	442..464	12/24 (50%) 22/24 (92%)	3e-07
zf-BED: domain 3 of 3	427..465	14/52 (27%) 27/52 (52%)	0.38
zf-C2H2: domain 12 of 13	470..492	12/24 (50%) 19/24 (79%)	0.00044
zf-C2H2: domain 13 of 13	498..520	12/24 (50%) 22/24 (92%)	9.8e-07

Example 118.

The NOV118 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 118A.

Table 118A. NOV118 Sequence Analysis		
	SEQ ID NO: 331	1899 bp
NOV118a, CG59805-01 DNA Sequence	CAAACTCTACTACCTCTATATGACATTTCAGGTGTCTGTGACCTTTGATGATGTGGCT GTGACTTTCACCCAGGAGGAGTGGGGCCAGCTGGACCTAGCTCAGCGGACCCTGTACC AGGAGGTGATGCTGGAAAACTGTGGGCTCCTGGTATCTCTGGGTGGGTGTCCTGTTCC CAGACCTGAGCTGATCTACCACCTAGAGCATGGGCAGGAGCCATGGACCAGGAAGGAA GACCTCTCCCAAGGCACCTGTCCAGGTGACAAAGGAAAACCCAAGAGCACAGAACCTA CCACCTGTGAGCTAGCCTTGTCTGAAGGAATCTCTTTTTGGGGACAACTAACACAAGG AGCTTCAGGGGACTCCCAGTTGGGGCAACCCAAGGATCAGGATGGGTTTTCAGAAATG CAGGGAGAACGCTTGAGACCAGGGTTAGATTCCCAAAAGGAGAAGCTTCCTGGAAAAA TGAGCCCCAAACATGATGGTTTAGGGACAGCTGATAGTGTGTGTTCAAGGATTATACA GGATCGAGTCTCCTTAGGAGATGATGTCCATGACTGTGACTCACATGGATCAGGTAAA AATCCAGTTATTCAGGAAGAGGAAAATATCTTTAAATGCAATGAATGTGAAAAAGTGT TTAACAAGAAACGCCTGCTTGCTCGGCATGAGAGGATTCACTCTGGAGTGAAGCCCTA TGAATGCACAGAGTGTGGAAAAACCTTTAGCAAGAGTACATACCTCCTGCAGCACCAC ATGGTCCACACTGGGGAGAAGCCCTATAAGTGCATGGAGTGTGGGAAGGCTTTTAATC GGAAGTCACACCTTACCCAGCACCAGCGGATTCACAGTGGAGAGAAGCCTTATAAGTG CAGTGAATGTGGAAAGGCCTTCACCCACCGCTCCACTTTTGTCTTGCATAACAGGAGC CACACTGGAGAAAAACCCTTTGTGTGCAAAGAGTGTGGCAAAGCCTTTCGAGATAGGC CAGGTTTCATTCGACACTACATCATCCACAGTGGTGAGAATCCCTACGAGTGCTTCGA ATGTGGCAAGGTCTTCAAACACAGATCATACCTCATGTGGCACCAGCAGACTCATACC GGGGAGAAGCCCTATGAGTGCAGTGAATGTGGGAAGGCCTTCTGTGAGAGCGCAGCGC TGATTCACCACTATGTCATCCACACTGGAGAGAAGCCCTTTGAGTGCCTCGAGTGTGG GAAGGCTTTCAACCACCGATCCTACCTCAAAAGGCACCAGCGGATTCACACTGGGGAG AAGCCATATGTGTGTAGTGAATGCGGAAAGGCCTTCACCCACTGCTCTACTTTCATCT TGCATAAAAGGGCCCACACTGGAGAAAAACCTTTCGAGTGCAAAGAGTGTGGGAAAGC CTTTAGCAATAGGGCAGACCTCATTCGCCACTTCAGCATCCACACTGGAGAGAAGCCC TATGAGTGCATGGAGTGTGGGAAAGGCCTTCAACCGCAGGTCAGGCCTCACAAGGCACC AGCGGATTCATAGTGGAGAGAAGCCCTATGAATGCATCGAGTGTGGGAAAACATTTTG CTGGAGCACAAACCTCATTCGACACTCTATCATCCACACTGGAGAGAAGCCGTATGAG TGCAGTGAATGTGGAAAGGCCTTCAGTCGCAGCTCGTCCCTCACTCAGCATCAAAGGA

450

	TGCATACTGGGAGAAATCCTATCAGTGTAACAGATGTGGGAAGACCTTTTACAAGTGG GCAGACCTCAGTCAACATCCAAGAACTTTTATTGGGGAAAAACTTTTTGAATGTCACC ACTGAGGAAAATCTTTTGCAAGAGGAAGCATCTTACATGGCATCTGATCGTACATACC AAAGAGAAACCCCACAAGTGTCTTCACTGTGAGAAAACCTTCT
	ORF Start: ATG at 20 ORF Stop: TGA at 1886
	SEQ ID NO: 332 | 622 aa | MW at 70677.2kD
NOV118a, CG59805-01 Protein Sequence	MTFQVSVTFDDVAVTFTQEEWGQLDLAQRTLYQEVMLENCGLLVSLGGCPVPRPELIY HLEHGQEPWTRKEDLSQGTCPGDKGKPKSTEPTTCELALSEGISFWGQLTQGASGDSQ LGQPKDQDGFSEMQGERLRPGLDSQKEKLPGKMSPKHDGLGTADSVCSRIIQDRVSLG DDVHDCDSHGSGKNPVIQEEENIFKCNECEKVFNKKRLLARHERIHSGVKPYECTECG KTFSKSTYLLQHHMVHTGEKPYKCMECGKAFNRKSHLTQHQRIHSGEKPYKCSECGKA FTHRSTFVLHNRSHTGEKPFVCKECGKAFRDRPGFIRHYIIHSGENPYECFECGKVFK HRSYLMWHQQTHTGEKPYECSECGKAFCESAALIHHYVIHTGEKPFECLECGKAFNHR SYLKRHQRIHTGEKPYVCSECGKAFTHCSTFILHKRAHTGEKPFECKECGKAFSNRAD LIRHFSIHTGEKPYECMECGKAFNRRSGLTRHQRIHSGEKPYECIECGKTFCWSTNLI RHSIIHTGEKPYECSECGKAFSRSSSLTQHQRMHTGRNPISVTDVGRPFTSGQTSVNI QELLLGKNFLNVTTEENLLQEEASYMASDRTYQRETPQVSSL

Further analysis of the NOV118a protein yielded the following properties shown in Table 118B.

Table 118B. Protein Sequence Properties NOV118a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3796 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV118a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded
5   several homologous proteins shown in Table 118C.

Table 118C. Geneseq Results for NOV118a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV118a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB22693	Protein #4692 encoded by probe for measuring heart cell gene expression - Homo sapiens, 468 aa. [WO200157274-A2, 09-AUG-2001]	81..548 1..468	468/468 (100%) 468/468 (100%)	0.0
AAM70526	Human bone marrow expressed probe encoded protein SEQ ID NO: 30832 - Homo sapiens, 468 aa. [WO200157276-A2, 09-AUG-2001]	81..548 1..468	468/468 (100%) 468/468 (100%)	0.0

AAM58080	Human brain expressed single exon probe encoded protein SEQ ID NO: 30185 - Homo sapiens, 468 aa. [WO200157275-A2, 09-AUG-2001]	81..548 1..468	468/468 (100%) 468/468 (100%)	0.0
AAM30843	Peptide #4880 encoded by probe for measuring placental gene expression - Homo sapiens, 468 aa. [WO200157272-A2, 09-AUG-2001]	81..548 1..468	468/468 (100%) 468/468 (100%)	0.0
AAM18364	Peptide #4798 encoded by probe for measuring cervical gene expression - Homo sapiens, 468 aa. [WO200157278-A2, 09-AUG-2001]	81..548 1..468	468/468 (100%) 468/468 (100%)	0.0

In a BLAST search of public sequence databases, the NOV118a protein was found to have homology to the proteins shown in the BLASTP data in Table 118D.

Table 118D. Public BLASTP Results for NOV118a				
Protein Accession Number	Protein/Organism/Length	NOV118a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O43296	Zinc finger protein 264 - Homo sapiens (Human), 627 aa.	4..622 11..627	530/619 (85%) 567/619 (90%)	0.0
Q96NL3	CDNA FLJ30663 FIS, CLONE FCBBF1000598, MODERATELY SIMILAR TO ZINC FINGER PROTEIN 84 - Homo sapiens (Human), 588 aa.	7..572 9..573	334/566 (59%) 403/566 (71%)	0.0
Q99676	Zinc finger protein 184 - Homo sapiens (Human), 751 aa.	2..571 23..623	280/604 (46%) 377/604 (62%)	e-160
P51523	Zinc finger protein 84 (Zinc finger protein HPF2) - Homo sapiens (Human), 738 aa.	4..617 5..626	286/637 (44%) 368/637 (56%)	e-157
Q9BX82	EZFIT-RELATED PROTEIN 1 - Homo sapiens (Human), 626 aa.	7..617 14..626	278/621 (44%) 364/621 (57%)	e-156

PFam analysis predicts that the NOV118a protein contains the domains shown in
5   the Table 118E.

Table 118E. Domain Analysis of NOV118a			
Pfam Domain	NOV118a Match Region	Identities/ Similarities for the Matched Region	Expect Value
KRAB: domain 1 of 1	7..70	41/66 (62%) 54/66 (82%)	2.2e-33
zf-C2H2: domain 1 of 13	198..220	11/24 (46%) 17/24 (71%)	3.9e-05
BolA: domain 1 of 1	161..238	14/88 (16%) 49/88 (56%)	3.4
zf-C2H2: domain 2 of 13·	226..248	10/24 (42%) 18/24 (75%)	6.2e-05
zf-C2H2: domain 3 of 13	254..276	14/24 (58%) 22/24 (92%)	5e-07
TFIIS: domain 1 of 1	257..292	12/39 (31%) 21/39 (54%)	5.7
zf-C2H2: domain 4 of 13	282..304	11/24 (46%) 20/24 (83%)	3.7e-05
LIM: domain 1 of 1	256..320	14/71 (20%) 48/71 (68%)	0.38
zf-C2H2: domain 5 of 13	310..332	8/24 (33%) 18/24 (75%)	7.6e-05
zf-C2H2: domain 6 of 13	338..360	11/24 (46%) 19/24 (79%)	1.1e-05
zf-C2H2: domain 7 of 13	366..388	9/24 (38%) 18/24 (75%)	0.00027
zf-C2H2: domain 8 of 13	394..416	12/24 (50%) 21/24 (88%)	7.9e-07
zf-C2H2: domain 9 of 13	422..444	10/24 (42%) 19/24 (79%)	0.00014
zf-C2H2: domain 10 of 13	450..472	10/24 (42%) 20/24 (83%)	8.3e-06
zf-C2H2: domain 11 of 13	478..500	13/24 (54%) 21/24 (88%)	3e-07
zf-BED: domain 1 of 1	463..501	14/52 (27%) 29/52 (56%)	0.1
zf-C2H2: domain 12 of 13	506..528	11/24 (46%) 17/24 (71%)	0.0016

zf-C2H2: domain 13 of 13	534..556	13/24 (54%) 23/24 (96%)	7.2e-08

Example 119.

The NOV119 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 119A.

Table 119A. NOV119 Sequence Analysis		
	SEQ ID NO: 333	1546 bp
NOV119a, CG59928-01 DNA Sequence	GCTCAGTAGGCGTCGGGCTGTGATGCCCCAACTGCTCCAGCGTCTGCAGGCGCGCGCG GGCGCGGTAGGCGTACTCGCTGGCCGGATAGCGCGTGATGATGAACTGGTAGGTCTGC GCCGCATCGACGAACAGGCTCTCGCGCTCCAGGCATTGACCGCGCAGCAGGGAAATCT CCGGCTGCAGGTAATTGCGTGAGCGGCTCTTGCGCTCGGCCTGCGACAGCTCCAGCGC GACACGGGCGCAATCGCCTTCGTTGTAGGCGCGATAGGCGTTGTTCAGATGATGGTCG AGCGAGACACGGGTGCAACCCGCAGCAACCAGGGCCACGGCCAGAATGATCAGGTTAC GCATGGGCAATTCCTCCAATGAGCAGTGTATCGACAGCCCAGGCAAAAACTGAACAGC GGCAAGCCGACGACGGTTTTTCTGGCGGCGGCCTTGGCATGACGCCACTGCCTCTCATT TTATCAACGCCAGCGCCACGACCGCTCGTCCTCTCGAACCAGCGCTAAATCCCCTTCT GCGCTGACCCATATCAATGCCGTTCAGCGCAACAGGGTGTGTAATGTAGGTACAGACT CCAGGCGAGGACGCTGCCATGAAACTGCAACGACTGTTGGTCGTCATCGACGCCGAAC ACCAGCAACAACCCGCCTGCAACGCGCAGCCGATGTGGCACGCAAGACCGGCGCCGA ATTGCACCTGTTGCAGATCGAATACCACCCAAGCCTGGAAAGCGGCCTGCTGGACAGC CATCTGCTCAACCGCGCCCGTGAAACCATCCTGCGACAGAGCCACGAGGCCCTGCGTG CCAGCGTCGCTCACCTGAGCGATGAAGGATTCAAGATCGCAGTGGACGTGCGCTGGGG CAAACGTCGTCATGAAGAAATCCTCGCCCGCGTCGCGGTGTTGCAACCGGACATCCTG TTCAAGTCGACTCATCCCAGCAGTGCGCTGCGCCGCCTGTTGTTCAGTGATACCAGTT GGCAGCTGATTCGCCGCAGCCCGGTGCCGCTGTGGCTGGTACACGACGCCGAGCCCCA TGGTCAGAGCCTGTGCGCTGCGCTCGACCCGCTGCACAGCGCGGACAAACCTGCCGCC CTCGATCATCAGTTGATTGATGCCAGCCAGACCCTGCAGGCCGAGCTCGGCTTACAGG CCCAATACCTGCATGCACAGGCGCCTCTGCCGCGGTCGCTGCTGTTCGACGCCGAGGT AGCGCAGGAATATGAAGACTACGTGACCCAGTGCAGCCGCGAGCACCGCGAAGCCTTC GACAAGCTGATCGCCCAGCACGCCATCGATAGAGCACAGGCCCACCTGTTGGACGGTT TTGCCGAGGAAGTCATCCCGCGTTTCGTGCGTGAGCACAATATAGGCCTGCTGGTGAT GGGCGCCATCGCCCGCGGCCATCTGGACAGCCTGCTGATCGGCCACACCGCAGAACGG GTGCTGGAACGTGTCGAGTGCGATCTGCTGGTGATCAAATCGCACGGCAAAGGGTAGT GCACAGGAACAATGACTACAGCCCGACGCCTACTGAGC
	ORF Start: ATG at 599	ORF Stop: TAG at 1505
	SEQ ID NO: 334	302 aa	MW at 33922.3kD
NOV119a, CG59928-01 Protein Sequence	MKLQRLLVVIDAEHQQQPALQRAADVARKTGAELHLLQIEYHPSLESGLLDSHLLNRA RETILRQSHEALRASVAHLSDEGFKIAVDVRWGKRRHEEILARVAVLQPDILFKSTHP SSALRRLLFSDTSWQLIRRSPVPLWLVHDAEPHGQSLCAALDPLHSADKPAALDHQLI DASQTLQAELGLQAQYLHAQAPLPRSLLFDAEVAQEYEDYVTQCSREHREAFDKLIAQ HAIDRAQAHLLDGFAEEVIPRFVREHNIGLLVMGAIARGHLDSLLIGHTAERVLERVE CDLLVIKSHGKG	

Further analysis of the NOV119a protein yielded the following properties shown in Table 119B.

Table 119B. Protein Sequence Properties NOV119a	
PSort analysis:	0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.2014 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

454

A search of the NOV119a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 119C.

Table 119C. Geneseq Results for NOV119a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV119a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found				

In a BLAST search of public sequence databases, the NOV119a protein was found
5   to have homology to the proteins shown in the BLASTP data in Table 119D.

Table 119D. Public BLASTP Results for NOV119a				
Protein Accession Number	Protein/Organism/Length	NOV119a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9HW73	HYPOTHETICAL PROTEIN PA4328 - Pseudomonas aeruginosa, 304 aa.	1..297 1..299	156/299 (52%) 200/299 (66%)	1e-79
Q9KS28	HYPOTHETICAL PROTEIN VC1433 - Vibrio cholerae, 315 aa.	5..300 6..304	78/302 (25%) 147/302 (47%)	4e-29
CAC91106	PUTATIVE STRESS PROTEIN - Yersinia pestis, 318 aa.	2..300 3..303	93/310 (30%) 137/310 (44%)	2e-28
AAL20579	PUTATIVE UNIVERSAL STRESS PROTEIN - Salmonella typhimurium LT2, 315 aa.	4..297 5..300	91/305 (29%) 139/305 (44%)	2e-28
CAD01669	CONSERVED HYPOTHETICAL PROTEIN - Salmonella enterica subsp. enterica serovar Typhi, 315 aa.	4..297 5..300	91/305 (29%) 139/305 (44%)	3e-28

PFam analysis predicts that the NOV119a protein contains the domains shown in the Table 119E.

455

<table>
<tr><td colspan="4" align="center">**Table 119E. Domain Analysis of NOV119a**</td></tr>
<tr><td>**Pfam Domain**</td><td>**NOV119a Match Region**</td><td>**Identities/ Similarities for the Matched Region**</td><td>**Expect Value**</td></tr>
<tr><td>Usp: domain 1 of 2</td><td>2..144</td><td>28/153 (18%) 92/153 (60%)</td><td>0.0014</td></tr>
<tr><td>Usp: domain 2 of 2</td><td>160..297</td><td>28/153 (18%) 88/153 (58%)</td><td>0.013</td></tr>
</table>

Example 120.

The NOV120 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 120A.

<table>
<tr><td colspan="3" align="center">**Table 120A. NOV120 Sequence Analysis**</td></tr>
<tr><td></td><td colspan="2">SEQ ID NO: 335 | 2202 bp</td></tr>
<tr><td>NOV120a, CG59947-01 DNA Sequence</td><td colspan="2">CACCCTCCCGCCCCGCCCCCCGTCCAATGCTGAGCTCAGTCTGCGTCTCGTCCTTCCG
CGGGCGCCAGGGGGCCAGCAAGCAGCAGCCGGCGCCACCGCCGCAGCCGCCCGAGGTC
CCCGGTGGCGACAGCGGCAAGATCGTGATCAACGTGGGCGGCGTGCGCCATGAGACGT
ACCGCTCGACGCTGCGCACCCTGCCGGGGACGCGGCTGGCCGGCCTGACGGAGCCCGA
GGCGGCGGCACGCTTCGACTACGACCCGGGCGCCGACGAGTTCTTCTTTGACCGGCAC
CCGGGAGTCTTCGCGTACGTGCTCAACTACTACCGCACCGGCAAGCTGCACTGCCCAG
CCGACGTGTGCGGGCCCCTGTTTGAGGAGGAGCTCGGCTTCTGGGGCATCGACGAGAC
CGACGTGGAGGCCTGCTGCTGGATGACCTACCGGCAGCATCGCGACGCTGAGGAGGCG
CTCGACTCCTTCGAGGCGCCCGACCCCGCGGGCGCCGCCAACGCCGCCAACGCCGCAG
GCGCCCACGACGGAGGCCTGGACGACGAGGCGGGCGCGGGCGGCGGCGGCCTGGACGG
AGCGGGCGGCGAGCTCAAGCGCCTCTGCTTCCAGGACGCGGGCGGCGGCGCCGGGGGG
CCGCCAGGGGGCGCGGGCGGCGCGGGCGGCACATGGTGGCGCCGCTGGCAGCCCCGCG
TGTGGGCGCTCTTCGAGGACCCCTACTCGTCGCGGGCTGCCAGGTATGTGGCCTTCGC
CTCCCTCTTCTTCATCCTCATCTCCATCACCACCTTCTGCCTGGAAACCCATGAGGGC
TTCATCCATATTAGCAACAAGACGGTGACCCAGGCCTCCCCGATCCCCGGGGCACCTC
CGGAGAACATCACCAACGTGGAGGTGGAGACGGAGCCCTTCCTGACCTACGTGGAGGG
GGTGTGCGTGGTCTGGTTCACCTTCGAGTTCCTCATGCGCATCACCTTCTGCCCAGAC
AAGGTGGAGTTTCTTAAAAGCAGCCTCAACATCATCGACTGTGTGGCCATCCTGCCCT
TCTATCTCGAGGTGGGCCTCTCGGGCCTCAGCTCCAAGGCCGCCAAAGACGTGCTGGG
CTTCCTGCGGGTGGTCCGCTTCGTCCGCATCCTGCGCATCTTCAAGCTGACCCGGCAC
TTCGTGGGGCTGCGCGTGCTGGGACACACGCTCCGCGCCAGCACCAACGAGTTCCTGC
TGCTCATCATCTTCCTGGCCCTGGGGGTGCTCATCTTCGCCACCATGATTTACTACGC
TGAGCGCATTGGCGCCGACCCCGATGACATCCTGGGCTCCAACCACACCTACTTCAAG
AACATCCCCATTGGCTTCTGGTGGGCTGTGGTCACCATGACGACCCTGGGCTATGGAG
ACATGTACCCCAAGACGTGGTCGGGGATGCTGGTCGGGGCGCTGTGTGCCCTGGCGGG
GGTGCTGACCATCGCCATGCCTGTGCCCGTCATTGTCAACAACTTTGGCATGTACTAT
TCGCTGGCCATGGCCAAGCAGAAGCTGCCCAAGAAGAAGAACAAACACATCCCCCGGC
CCCCGCAACCGGGCTCGCCCAACTACTGCAAGCCTGACCCACCCCCGCCACCCCCGCC
CCACCCGCACCACGGCAGCGGGGGCATCAGCCCGCCCACCCATCACCCCACCCTCC
ATGGGGGTGACTGTGGCCGGGGCCTACCCAGCGGGGCCCCACACGCACCCCGGGCTGC
TCAGGGGGGGAGCGGGTGGGCTGGGGATCATGGGGCTGCCTCCTCTGCCAGCCCCCGG
CGAGCCTTGCCCGTTGGCTCAGGAGGAGGTGATTGAGATCAACCGGGCAGATCCTCGC
CCCAATGGGGATCCGGCAGCAGCTGCGCTTGCCCACGAGGACTGCCCAGCCATTGACC
AGCCTGCCATGTCCCCGGAAGACAAGAGCCCCATCACGCCTGGAAGCCGTGGCCGCTA
TAGCCGGGACCGAGCCTGCTTCCTCCTCACCGACTATGCCCCTTCCCCTGATGGCTCC
ATCCGAAAAGCCACTGGTGCTCCCCCACTGCCCCCCCAAGACTGGCGTAAGCCAGGCC
CCCCAAGCTTCTTGCCCGACCTCAACGCCAACGCCGCGGCCTGGATATCCCCCTAGTG
GACGAACCCCCTCCCCCCGGGCTCTTGTCACCGCCTGAGACCTCGCGAGACTTTCG</td></tr>
<tr><td></td><td colspan="2">ORF Start: ATG at 27 | ORF Stop: TAG at 2142</td></tr>
<tr><td></td><td colspan="2">SEQ ID NO: 336 | 705 aa | MW at 75590.5kD</td></tr>
<tr><td>NOV120a,</td><td colspan="2">MLSSVCVSSFRGRQGASKQQPAPPPQPPEVPGGDSGKIVINVGGVRHETYRSTLRTLP
GTRLAGLTEPEAAARFDYDPGADEFFFDRHPGVFAYVLNYYRTGKLHCPADVCGPLFE</td></tr>
</table>

456

| CG59947-01 Protein Sequence | EELGFWGIDETDVEACCWMTYRQHRDAEEALDSFEAPDPAGAANAANAAGAHDGGLDD EAGAGGGGLDGAGGELKRLCFQDAGGGAGGPPGGAGGAGGTWWRRWQPRVWALFEDPY SSRAARYVAFASLFFILISITTFCLETHEGFIHISNKTVTQASPIPGAPPENITNVEV ETEPFLTYVEGVCVVWFTFEFLMRITFCPDKVEFLKSSLNIIDCVAILPFYLEVGLSG LSSKAAKDVLGFLRVVRFVRILRIFKLTRHFVGLRVLGHTLRASTNEFLLLIIFLALG VLIFATMIYYAERIGADPDDILGSNHTYFKNIPIGFWWAVVTMTTLGYGDMYPKTWSG MLVGALCALAGVLTIAMPVPVIVNNFGMYYSLAMAKQKLPKKKNKHIPRPPQPGSPNY CKPDPPPPPPPHPHHGSGGISPPPPITPPSMGVTVAGAYPAGPHTHPGLLRGGAGGLG IMGLPPLPAPGEPCPLAQEEVIEINRADPRPNGDPAAAALAHEDCPAIDQPAMSPEDK SPITPGSRGRYSRDRACFLLTDYAPSPDGSIRKATGAPPLPPQDWRKPGPPSFLPDLN ANAAAWISP |

Further analysis of the NOV120a protein yielded the following properties shown in Table 120B.

Table 120B. Protein Sequence Properties NOV120a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.5071 probability located in mitochondrial inner membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV120a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 120C.

Table 120C. Geneseq Results for NOV120a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV120a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY34120	Human potassium channel K+Hnov4 - Homo sapiens, 601 aa. [WO9943696-A1, 02-SEP-1999]	32..526 4..476	371/510 (72%) 399/510 (77%)	0.0
AAY32016	Caenorhabditis elegans cation channel protein - Caenorhabditis elegans, 556 aa. [WO9947923-A2, 23-SEP-1999]	33..512 27..465	217/486 (44%) 300/486 (61%)	e-113
AAB86319	Human Kv4.2 protein - Homo sapiens, 629 aa. [DE19963612-A1, 12-JUL-2001]	16..521 22..441	173/511 (33%) 256/511 (49%)	5e-69
AAY13523	Amino acid sequence of KV4.2FL ion channel protein - Mammalia, 630 aa. [WO9923880-A1, 20-MAY-1999]	16..521 23..442	173/511 (33%) 257/511 (49%)	8e-68

AAW42996	Putative mature potassium channel 2 protein - Homo sapiens, 494 aa. [US5710019-A, 20-JAN-1998]	17..510 4..425	171/503 (33%) 240/503 (46%)	2e-66

In a BLAST search of public sequence databases, the NOV120a protein was found to have homology to the proteins shown in the BLASTP data in Table 120D.

Table 120D. Public BLASTP Results for NOV120a

Protein Accession Number	Protein/Organism/Length	NOV120a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q14003	Voltage-gated potassium channel protein Kv3.3 (KSHIIID) - Homo sapiens (Human), 757 aa.	1..705 1..757	704/757 (92%) 704/757 (92%)	0.0
Q01956	Voltage-gated potassium channel protein Kv3.3 (KSHIIID) - Rattus norvegicus (Rat), 889 aa.	1..693 1..756	663/757 (87%) 668/757 (87%)	0.0
Q63959	Voltage-gated potassium channel protein Kv3.3 (KSHIIID) - Mus musculus (Mouse), 769 aa.	1..671 1..724	650/725 (89%) 653/725 (89%)	0.0
A42073	potassium channel protein Kv3.3 - mouse, 679 aa.	32..607 8..581	557/576 (96%) 559/576 (96%)	0.0
Q9PVD1	KV3.1 POTASSIUM CHANNEL - Xenopus laevis (African clawed frog), 592 aa.	34..671 6..547	441/640 (68%) 479/640 (73%)	0.0

PFam analysis predicts that the NOV120a protein contains the domains shown in the Table 120E.

Table 120E. Domain Analysis of NOV120a

Pfam Domain	NOV120a Match Region	Identities/ Similarities for the Matched Region	Expect Value
K_tetra: domain 1 of 1	36..137	50/112 (45%) 86/112 (77%)	1.6e-47
thaumatin: domain 1 of 1	314..319	4/6 (67%) 6/6 (100%)	0.7

ion_trans: domain 1 of 1	295..486	51/231 (22%) 155/231 (67%)	2.1e-29

Example 121.

The NOV121 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 121A.

Table 121A. NOV121 Sequence Analysis		
	SEQ ID NO: 337	1943 bp
NOV121a, CG59938-01 DNA Sequence	AGATCCACGTGATCTCCAAAGACCCCTGTTGTGTTGTGTTGGGAGGTGGATCCTGAAT CCACCCAGAGAAGCCTGATACCAATAAAATCCCTGCTTGCTTTCCAGGAGACCCTTGG TCTTCATGTCTTTGGTGTGTGCACTCTTGAACACATGCCAGGCACACAGGGTGCATGA CGACAAGCCTAATATTGTCCTAATCATGGTTGATGACCTGGGTATTGGAGATCTGGGC TGCTACGGCAATGACACCATGAGGACGCCTCACATCGACCGCCTTGCCAGGGAAGGCG TGCGACTGACTCAGCACATCTCTGCCGCCTCCCTCTGCAGCCCAAGCCGGTCCGCGTT CTTGACGGGAAGATACCCCATCCGATCAGGTATGGTTTCTAGTGGTAATAGACGTGTC ATCCAAAATCTTGCAGTCCCCGCAGGCCTCCCTCTTAATGAGACAACACTTGCAGCCT TGCTAAAGAAGCAAGGATACAGCACGGGGCTTATAGGTAAGTTAGGCAAATGGCACCT GGGTTTGAGCTGCGCCTCTCGGAATGATCACTGTTACCACCCGCTCAACCATGGTTTT CACTACTTTTACGGGGTGCCTTTTGGACTTTTAAGCGACTGCCAGGCATCCAAGACAC CAGAACTGCACCGCTGGCTCAGGATCAAACTGTGGATCTCCACGGTAGCCCTTGCCCT GGTTCCTTTTCTGCTTCTCATTCCCAAGTTCGCCCGCTGGTTCTCAGTGCCATGGAAG GTCATCTTTGTCTTTGCTCTCCTCGCCTTTCTGTTTTTCACTTCCTGGTACTCTAGTT ATGGATTTACTCGACGTTGGAATTGCATCCTTATGAGGAACCATGAAATTATCCAGCA GCCAATGAAAGAGGAGAAAGTAGCTTCCCTCATGCTGAAGGAGGCACTTGCTTTCATT GAAAGGTACAAAAGGGAACCTTTTCTCCTCTTTTTTTCCTTCCTGCACGTACATACTC CACTCATCTCCAAAAAGAAGTTTGTTGGGCGCAGTAAATATGGCAGGTATGGGGACAA TGTAGAAGAAATGGATTGGATGGTGGGTGGTAAAATCCTGGATGCCCTGGACCAGGAG CGCCTGGCCAACCACACCTTGGTGTACTTCACCTCTGACAACGGGGGCCACCTGGAGC CCCTGGACGGGGCTGTTCAGCTGGGTGGCTGGAACGGGATCTACAAAGGTGGCAAAGG AATGGGAGGATGGGAAGGAGGTATCCGTGTGCCAGGGATATTCCGGTGGCCGTCAGTC TTGGAGGCTGGGAGAGTGATCAATGAGCCCACCAGCTTAATGGACATCTATCCGACGC TGTCTTATATAGGCGGAGGGATCTTGTCCCAGGACAGAGTGATTGACGGCCAGAACCT AATGCCCCTGCTGGAAGGAAGGGCGTCCCACTCCGACCACGAGTTCCTCTTCCACTAC TGTGGGGTCTATCTGCACACGGTCAGGTGGCATCAGAAGGACACTGTGTGGAAAGCTC ATTATGTGACTCCTAAATTCTACCCTGAAGGAACAGGTGCCTGCTATGGGAGTGGAAT ATGTTCATGTTCGGGGGATGTAACCTACCACGACCCACCACTCCTCTTTGACATCTCA AGAGACCCTTCAGAAGCCCTTCCACTGAACCCTGACAATGAGCCATTATTTGACTCCG TGATCAAAAAGATGGAGGCAGCCATAAGAGAGCATCGTAGGACACTAACACCTGTCCC ACAGCAGTTCTCTGTGTTCAACACAATTTGGAAACCATGGCTGCAGCCTTGCTGTGGG ACCTTCCCCTTCGTGGGTGTGACAAGGAAGATGACATCCTTCCCATGGCTCCCTGAG ACCATGCGGACCACGTGTTACCCACCACAAACTTACTGTTACAATGGTCATAGGAGCA GAGCTCACCTGACTGATTCATTCCATTTG
	ORF Start: ATG at 122	ORF Stop: TGA at 1853
	SEQ ID NO: 338	577 aa	MW at 65099.5kD
NOV121a, CG59938-01 Protein Sequence	MSLVCALLNTCQAHRVHDDKPNIVLIMVDDLGIGDLGCYGNDTMRTPHIDRLAREGVR LTQHISAASLCSPSRSAFLTGRYPIRSGMVSSGNRRVIQNLAVPAGLPLNETTLAALL KKQGYSTGLIGKLGKWHLGLSCASRNDHCYHPLNHGFHYFYGVPFGLLSDCQASKTPE LHRWLRIKLWISTVALALVPFLLLIPKFARWFSVPWKVIFVFALLAFLFFTSWYSSYG FTRRWNCILMRNHEIIQQPMKEEKVASLMLKEALAFIERYKREPFLLFFSFLHVHTPL ISKKKFVGRSKYGRYGDNVEEMDWMVGGKILDALDQERLANHTLVYFTSDNGGHLEPL DGAVQLGGWNGIYKGGKGMGGWEGGIRVPGIFRWPSVLEAGRVINEPTSLMDIYPTLS YIGGGILSQDRVIDGQNLMPLLEGRASHSDHEFLFHYCGVYLHTVRWHQKDTVWKAHY VTPKFYPEGTGACYGSGICSCSGDVTYHDPPLLFDISRDPSEALPLNPDNEPLFDSVI KKMEAAIREHRRTLTPVPQQFSVFNTIWKPWLQPCCGTFPFCGCDKEDDILPMAP

Further analysis of the NOV121a protein yielded the following properties shown in Table 121B.

Table 121B. Protein Sequence Properties NOV121a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV121a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 121C.

Table 121C. Geneseq Results for NOV121a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV121a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM78688	Human protein SEQ ID NO 1350 - Homo sapiens, 590 aa. [WO200157190-A2, 09-AUG-2001]	1..572 10..580	388/576 (67%) 449/576 (77%)	0.0
AAM39343	Human polypeptide SEQ ID NO 2488 - Homo sapiens, 589 aa. [WO200153312-A1, 26-JUL-2001]	20..571 37..587	331/555 (59%) 404/555 (72%)	0.0
AAM41129	Human polypeptide SEQ ID NO 6060 - Homo sapiens, 646 aa. [WO200153312-A1, 26-JUL-2001]	20..571 94..644	331/555 (59%) 404/555 (72%)	0.0
AAY39920	Human steroid sulphatase protein sequence - Homo sapiens, 583 aa. [WO9950453-A1, 07-OCT-1999]	20..569 26..575	295/559 (52%) 374/559 (66%)	e-166
AAB51185	Human sulfatase protein C SEQ ID NO:14 - Homo sapiens, 583 aa. [US6153188-A, 28-NOV-2000]	20..569 26..575	294/559 (52%) 372/559 (65%)	e-165

5   In a BLAST search of public sequence databases, the NOV121a protein was found to have homology to the proteins shown in the BLASTP data in Table 121D.

Table 121D. Public BLASTP Results for NOV121a				
Protein Accession Number	Protein/Organism/Length	NOV121a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P54793	Arylsulfatase F precursor (EC 3.1.6.-) (ASF) - Homo sapiens (Human), 591 aa.	1..572 10..581	379/577 (65%) 441/577 (75%)	0.0
AAH20229	HYPOTHETICAL 64.9 KDA PROTEIN - Homo sapiens (Human), 593 aa.	4..574 24..593	358/574 (62%) 440/574 (76%)	0.0
P51689	Arylsulfatase D precursor (EC 3.1.6.-) (ASD) - Homo sapiens (Human), 593 aa.	4..574 24..593	349/574 (60%) 429/574 (73%)	0.0
P51690	Arylsulfatase E precursor (EC 3.1.6.-) (ASE) - Homo sapiens (Human), 589 aa.	20..571 37..587	334/555 (60%) 405/555 (72%)	0.0
P08842	Steryl-sulfatase precursor (EC 3.1.6.2) (Steroid sulfatase) (Steryl-sulfate sulfohydrolase) (Arylsulfatase C) (ASC) - Homo sapiens (Human), 583 aa.	20..569 26..575	295/559 (52%) 374/559 (66%)	e-166

PFam analysis predicts that the NOV121a protein contains the domains shown in the Table 121E.

Table 121E. Domain Analysis of NOV121a			
Pfam Domain	NOV121a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Sulfatase: domain 1 of 1	21..504	231/530 (44%) 410/530 (77%)	1e-187

Example 122.

5   The NOV122 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 122A.

461

Table 122A. NOV122 Sequence Analysis		
	SEQ ID NO: 339	3005 bp
NOV122a, CG59746-01 DNA Sequence	ATTTCTTTGGTGTTGTCTTCACAGCTGAACTTGCAAAACAGATTGGAACTTCAAGATT ATCAATAATCGGAGATACGTATATTTTATTTGTAAAGAAAAC**ATG**GCTGCCCTATTCC TACGTGGTTTTGTCCAAATAGGGAACTGCAAGACTGGGATATCTAAGTCAAAAGAAGC ATTCATTGAAGCAGTGGAAAGAAAGAAGAAAGATAGACTGGTGCTGTATTTCAAAAGT GGAAAATATAGCACTTTTCGGCTAAGTGATAATATTCAAAATGTAGTCCTTAAATCCT ATAGAGGAAACCAAAATCACCTGCATTTAACTTTACAAAATAATAATGGCTTGTTTAT TGAAGGATTATCCTCCACAGATGCTGAACAATTGAAGATATTCTTGGACAGAGTTCAT CAAAACGAGGTTCAGCCACCTGTGAGACCTGGTAAGGGTGGGAGTGTCTTTTCTAGCA CAACACAGAAGGAAATCAACAAAACTTCATTCCACAAAGTTGATGAGAAATCAAGTAG CAAATCTTTTGAGATAGCAAAAGGAAGTGGGACAGGTGTCCTTCAGAGGATGCCTTTG CTTACATCAAAATTGACACTTACTTGCGGAGAGTTATCAGAAAATCAGCACAAGAAGA GGAAAAGAATGCTCTCATCTAGCTCAGAGATGAATGAGGAATTCTTGAAAGAAAATAA TTCTGTAGAATACAAGAAATCCAAGGCAGATTGTTCGAGGTGTGTAAGCTATAATCGA GAGAAACAATTGAAGTTAAAAGAGTTAGAAGAGAATAAGAAATTGGAATGTGAATCTT CATGCATCATGAACGCCACTGGAAATCCTTACCTAGATGACATTGGTCTTCTCCAAGC TCTCACTGAGAAAATGGTTTTGGTATTTCTGTTACAACAAGGGTATAGTGACGGTTAC ACAAAGTGGGATAAATTAAAACTATTTTTTGAATTATTTCCAGAGAAAATATGCCACG GCCTCCCCAATTTGGGAAACACCTGTTATATGAATGCAGTGTTACAGTCTCTACTTTC AATCCCATCGTTTGCTGATGATTTACTTAATCAGAGTTTCCCATGGGGTAAAATTCCC CTTAATGCTCTTACCATGTGCTTGGCACGGCTACTTTTTTTTAAAGATACCTATAATA TAGAAATCAAGGAGATGTTACTCTTGAATCTTAAAAAGGCCATTTCAGCAGCTGCAGA GATATTCCATGGCAATGCACAGAACGATGCTCATGAGTTTTTAGCTCACTGTTTAGAT CAACTGAAAGATAACATGGAAAAACTCAACACAATTTGGAAGCCTAAAAGTGAATTTG GGGAAGATAATTTTCCTAAACAGGTTTTTGCTGATGATCCTGACACCAGTGGGTTTTC TTGCCCTGTCATTACTAATTTTGAGTTAGAGTTGTTGCACCTCCATTGCTTGTAAAGCT TGTGGTCAGGTTATTCTCAAGACAGAACTGAATAATTACCTCTCCATCAACCTTCCCC AAAGAATAAAAGCACATCCTTCATCTATTCAGTCTACTTTTGATCTTTTTTTTGGAGC AGAAGAGCTTGAGTATAAATGTGCAAAATGTGAGCACAAGACTTCCGTTGGAGTGCAC TCATTCAGTAGGCTACCTAGAATCCTTATTGTTCACCTCAAACGCTATAGCTTGAATG AGTTTTGTGCATTAAAGAAGAATGACCAGGAAGTCATCATTTCCAAATATTTAAAGGT GTCTTCTCATTGCAATGAAGGCACCAGACCACCTCTTCCCTTGAGTGAGGATGGAGAA ATTACAGATTTCCAATTATTAAAAGTTATTCGAAAGATGACTTCTGGAAACATCAGTG TATCATGGCCTGCAACAAAGGAATCCAAAGATATCCTGGCTCCACACATTGGATCAGA TAAGGAGTCTGAACAAAAAAAAGGCCAGACAGTCTTTAAAGGGGGCAAGCAGAAGACAG CAGCAAAAGTACCTTGGAAAAAATTCTAAACCAAATGAGCTAGAATCTGTATACTCAG GAGATCGAGCATTCATTGAAAAAGAACCGTTAGCTCACTTAATGACGTATCTGGAAGA TACCTCACTTTGTCAGTTCCACAAAGCTGGAGGTAAACCTGCCAGCAGCCCAGGCACA CCTCTCTCAAAAGTTGACTTTCAAACAGTGCCCGAAAATCCAAAACGAAAGAAATATG TGAAAACCAGTAAGTTTGTAGCTTTTGATAGGATTATCAATCCTACTAAAGATTTGTA TGAAGATAAAAATATCAGAATTCCAGAAAGATTCCAAAAAGTGTCTGAACAGACTCAG CAGTGTGACGGTATGAGAATCTGTGAACAAGCCCCTCAGCAGGCACTGCCTCAAAGCT TTCCAAAGCCAGGCACCCAGGGGCACACAAAGAACCTCCTAAGACCTACAAAATTAAA TCTACAGAAGTCTAACAGGAATTCCCTACTTGCACTGGGTTCCAATAAGAATCCAAGA AACAAAGACATTTTAGATAAGATAAAATCTAAAGCCAAGGAAACAAAAAGAAATGATG ATAAGGGAGATCATACCTACCGGCTCATTAGTGTTGTCAGCCATCTTGGGAAGACTCT AAAGTCAGGCCATTATATCTGTGATGCCTATGACTTTGAGAAACAGATCTGGTTCACT TACGATGATATGCGGGTGTTAGGTATCCAGGAGGCCCAGATGCAGGAGGATAGGCGTT GCACTGGGTACATCTTCTTTTACATGCATAATGAGATCTTTGAAGAGATGTTGAAAAG AGAAGAGAATGCCCAGCTTAATAGCAAGGAGGTAGAGGAGACCCTTCAGAAGGA**ATAA** GAGGAACGTACTCCTCCTTGTACAGATCTGCCTGACTGTCTCACTCGATACCACTTCC TCCATGGAAGGAAAACTGTGAACTTTATCCAGAGATGAAAATGCAATTAGTCTAGGAC CAAAGGTCAAACAGAAACACTTAATGGGGAGATCTGCATTCTAATCC
	ORF Start: ATG at 101	ORF Stop: TAA at 2840
	SEQ ID NO: 340	913 aa	MW at 104046.0kD
NOV122a, CG59746-01 Protein Sequence	MAALFLRGFVQIGNCKTGISKSKEAFIEAVERKKKDRLVLYFKSGKYSTFRLSDNIQN VVLKSYRGNQNHLHLTLQNNNGLFIEGLSSTDAEQLKIFLDRVHQNEVQPPVRPGKGG SVFSSTTQKEINKTSFHKVDEKSSSKSFEIAKGSGTGVLQRMPLLTSKLTLTCGELSE NQHKKRKRMLSSSSEMNEEFLKENNSVEYKKSKADCSRCVSYNREKQLKLKELEENKK LECESSCIMNATGNPYLDDIGLLQALTEKMVLVFLLQQGYSDGYTKWDKLKLFFELFP EKICHGLPNLGNTCYMNAVLQSLLSIPSFADDLLNQSFPWGKIPLNALTMCLARLLFF KDTYNIEIKEMLLLNLKKAISAAAEIFHGNAQNDAHEFLAHCLDQLKDNMEKLNTIWK PKSEFGEDNFPKQVFADDPDTSGFSCPVITNFELELLHSIACKACGQVILKTELNNYL SINLPQRIKAHPSSIQSTFDLFFGAEELEYKCAKCEHKTSVGVHSFSRLPRILIVHLK RYSLNEFCALKKNDQEVIISKYLKVSSHCNEGTRPPLPLSEDGEITDFQLLKVIRKMT SGNISVSWPATKESKDILAPHIGSDKESEQKKGQTVFKGASRRQQQKYLGKNSKPNEL		

462

	ESVYSGDRAFIEKEPLAHLMTYLEDTSLCQFHKAGGKPASSPGTPLSKVDFQTVPENP KRKKYVKTSKFVAFDRIINPTKDLYEDKNIRIPERFQKVSEQTQQCDGMRICEQAPQQ ALPQSFPKPGTQGHTKNLLRPTKLNLQKSNRNSLLALGSNKNPRNKDILDKIKSKAKE TKRNDDKGDHTYRLISVVSHLGKTLKSGHYICDAYDFEKQIWFTYDDMRVLGIQEAQM QEDRRCTGYIFFYMHNEIFEEMLKREENAQLNSKEVEETLQKE

Further analysis of the NOV122a protein yielded the following properties shown in Table 122B.

Table 122B. Protein Sequence Properties NOV122a	
PSort analysis:	0.7000 probability located in nucleus; 0.4270 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1047 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 16 and 17

A search of the NOV122a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 122C.

Table 122C. Geneseq Results for NOV122a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV122a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU07888	Polypeptide sequence for human hspG25 - Homo sapiens, 913 aa. [WO200166752-A2, 13-SEP-2001]	1..913 1..913	913/913 (100%) 913/913 (100%)	0.0
AAB75607	Human cancer associated antigen precursor HOM-TES-84/6 SEQ ID NO:6 - Homo sapiens, 912 aa. [WO200100874-A2, 04-JAN-2001]	1..905 1..904	429/920 (46%) 566/920 (60%)	0.0
AAU07869	Polypeptide sequence for mammalian Spg25 - Mammalia, 835 aa. [WO200166752-A2, 13-SEP-2001]	1..904 1..834	335/921 (36%) 504/921 (54%)	e-147
AAG75460	Human colon cancer antigen protein SEQ ID NO:6224 - Homo sapiens, 109 aa. [WO200122920-A2, 05-APR-2001]	810..912 3..107	61/105 (58%) 79/105 (75%)	3e-28
AAB39364	Gene 8 human secreted protein homologous amino acid sequence #113 - Bos taurus, 64 aa. [WO200057903-A2, 05-OCT-2000]	810..871 1..64	39/64 (60%) 48/64 (74%)	5e-15

463

In a BLAST search of public sequence databases, the NOV122a protein was found to have homology to the proteins shown in the BLASTP data in Table 122D.

Table 122D. Public BLASTP Results for NOV122a				
Protein Accession Number	Protein/Organism/Length	NOV122a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BXU7	Ubiquitin carboxyl-terminal hydrolase 26 (EC 3.1.2.15) (Ubiquitin thiolesterase 26) (Ubiquitin-specific processing protease 26) (Deubiquitinating enzyme 26) - Homo sapiens (Human), 913 aa.	1..913 1..913	913/913 (100%) 913/913 (100%)	0.0
Q9HBJ7	UBIQUITIN-SPECIFIC PROCESSING PROTEASE - Homo sapiens (Human), 922 aa.	1..905 1..904	429/920 (46%) 566/920 (60%)	0.0
Q9HCH8	KIAA1594 PROTEIN - Homo sapiens (Human), 931 aa (fragment).	50..912 3..929	393/932 (42%) 535/932 (57%)	e-171
Q99MX1	Ubiquitin carboxyl-terminal hydrolase 26 (EC 3.1.2.15) (Ubiquitin thiolesterase 26) (Ubiquitin-specific processing protease 26) (Deubiquitinating enzyme 26) - Mus musculus (Mouse), 835 aa.	1..904 1..834	335/921 (36%) 504/921 (54%)	e-147
Q9ES63	UBIQUITIN-SPECIFIC PROCESSING PROTEASE - Mus musculus (Mouse), 869 aa.	1..908 1..848	341/933 (36%) 480/933 (50%)	e-131

PFam analysis predicts that the NOV122a protein contains the domains shown in the Table 122E.

Table 122E. Domain Analysis of NOV122a			
Pfam Domain	NOV122a Match Region	Identities/ Similarities for the Matched Region	Expect Value
UCH-1: domain 1 of 1	295..326	21/32 (66%) 29/32 (91%)	8.8e-12
UCH-2: domain 1 of 1	820..885	20/72 (28%) 47/72 (65%)	2.2e-11

Example 123.

The NOV123 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 123A.

Table 123A. NOV123 Sequence Analysis			
	SEQ ID NO: 341	2146 bp	
NOV123a, CG88613-01 DNA Sequence	GAAGGAGCGGGCATGAGGCGCTGCCCGTGCCGTGGGAGCCTGAACGAGGCGGAGGCCG GGGCGCTGCCCGCGGCGGCCCGCATGGGACTGGAGGCGCCGCGAGGAGGGCGGCGGCG GCAGCCGGGACAGCAGCGACCTGGGCCCGGCGCAGGGGCCCCGGCGGGGCGGCCGGAG GGGGGCGGGCCCTGGGCCCGGACAGAGGGGTCCAGCCTCCACAGCGAGCCTGAGAGGG CCGGCCTCGGGCCTGCGCCGGGGACAGAGAGTCCGCAGGCAGAATTCTGGACAGACGG ACAGACTGAGCCCGCGGCAGCTGGCCTTGGAGTAGAGACCGAGAGGCCCAAGCAAAAG ACGGAGCCAGACAGGTCCAGCCTCCGGACGCATCTAGAATGGAGCTGGTCAGAGCTGG AGACGACTTGTCTTTGGACGGAGACCGGGACAGATGGCCTTTGGACTGATCCGCACAG GTCCGACCTCCAGTTTCAGCCCGAGGAGGCCAGCCCCTGGACACAGCCAGGGGTTCAT GGGCCCTGGACAGAGCTGGAAACGCATGGGTCACAGACTCAGCCAGAGAGGGTCAAGT CCTGGGCTGATAACCTCTGGACCCACCAGAACAGTTCCAGCCTCCAGACTCACCCAGA AGGAGCCTGTCCCTCAAAAGAGCCAAGTGCTGATGGCTCCTGGAAAGAATTGTATACT GATGGCTCCAGGACACAACAGGATATTGAAGGTCCCTGGACAGAGCCATATACTGATG GCTCCCAGAAAAAACAGGATACTGAAGCAGCCAGGAAACAGCCTGGCACTGGTGGTTT CCAAATACAACAGGATACTGATGGCTCCTGGACACAACCTAGCACTGACGGTTCCCAG ACAGCACCTGGGACAGACTGCCTCTTGGGAGAGCCTGAGGATGGCCCATTAGAGGAAC CAGAGCCTGGAGAATTGCTGACTCACCTGTACTCTCACCTGAAGTGTAGCCCCCTGTG CCCTGTGCCCCGCCTCATCATTACCCCTGAGACCCCTGAGCCTGAGGCCCAGCCAGTG GGACCCCCCTCCCGGGTTGAGGGGGGCAGCGGCGGCTTCTCCTCTGCCTCTTCTTTCG ACGAGTCTGAGGATGACGTGGTGGCCGGGGGCGGAGGTGCCAGCGATCCCGAGGACAG GTCTGGGAGCAAACCCTGGAAGAAGCTGAAGACAGTTCTGAAGTATTCACCCTTTGTG GTCTCCTTCCGAAAACACTACCCTTGGGTCCAGCTTTCTGGACATGCTGGGAACTTCC AGGCAGGAGAGGATGGTCGGATTCTGAAACGTTTCTGTCAGTGTGAGCAGCGCAGCCT GGAGCAGCTGATGAAAGACCCGCTGCCACCTTTCGTGCCTGCCTACTATGGCATGGTG CTGCAGGATGGCCAGACCTTCAACCAGATGGAAGACCTCCTGGCTGACTTTGAGGGCC CCTCCATTATGGACTGCAAGATGGGCAGCAGGACCTATCTGGAAGAGGAGCTAGTGAA GGCACGGGAACGTCCCCGTCCCCGGAAGGACATGTATGAGAAGATGGTGGCTGTGGAC CCTGGGGCCCCTACCCCTGAGGAGCATGCCCAGGGTGCAGTCACCAAGCCCCGCTACA TGCAGTGGAGGGAAACCATGAGCTCCACCTCTACCCTGGGCTTCCGGATCGAGGGCAT CAAGAAGGCAGATGGGACCTGTAACACCAACTTCAAGAAGACGCAGGCACTGGAGCAG GTGACAAAAGTGCTGGAGGACTTCGTGGATGGAGACCACGTCATCCTGCAAAAGTACG TGGCATGCCTAGAAGAACTTCGTGAAGCTCTGGAGATCTCCCCCTTCTTCAAGACCCA CGAGGTGGTAGGCAGCTCCCTCCTCTTCGTGCACGACCACACCGGCCTGGCCAAGGTC TGGATGATAGACTTCGGCAAGACGGTGGCCTTGCCCGACCACCAGACGCTCAGCCACA GGCTGCCCTGGGCTGAGGGCAACCGTGAGGACGGCTACCTCTGGGGCCTGGACAACAT GATCTGCCTCCTGCAGGGGCTGGCACAGAGCTGAGCTGCTCAGCCACCATCAGGTTAA TTGGATGGCGCCAGTCTGGCTGGAGGAGCCCTGAGATGCCATGGGAGGCCTGAGGTTG		
	ORF Start: ATG at 13	ORF Stop: TGA at 2062	
	SEQ ID NO: 342	683 aa	MW at 75206.8kD
NOV123a, CG88613-01 Protein Sequence	MRRCPCRGSLNEAEAGALPAAARMGLEAPRGGRRQPGQQRPGPGAGAPAGRPEGGGP WARTEGSSLHSEPERAGLGPAPGTESPQAEFWTDGQTEPAAAGLGVETERPKQKTEPD RSSLRTHLEWSWSELETTCLWTETGTDGLWTDPHRSDLQFQPEEASPWTQPGVHGPWT ELETHGSQTQPERVKSWADNLWTHQNSSSLQTHPEGACPSKEPSADGSWKELYTDGSR TQQDIEGPWTEPYTDGSQKKQDTEAARKQPGTGGFQIQQDTDGSWTQPSTDGSQTAPG TDCLLGEPEDGPLEEPEPGELLTHLYSHLKCSPLCPVPRLIITPETPEPEAQPVGPPS RVEGGSGGFSSASSFDESEDDVVAGGGGASDPEDRSGSKPWKKLKTVLKYSPFVVSFR KHYPWVQLSGHAGNFQAGEDGRILKRFCQCEQRSLEQLMKDPLRPFVPAYYGMVLQDG QTFNQMEDLLADFEGPSIMDCKMGSRTYLEEELVKARERPRPRKDMYEKMVAVDPGAP TPEEHAQGAVTKPRYMQWRETMSSTSTLGFRIEGIKKADGTCNTNFKKTQALEQVTKV LEDFVDGDHVILQKYVACLEELREALEISPFFKTHEVVGSSLLFVHDHTGLAKVWMID FGKTVALPDHQTLSHRLPWAEGNREDGYLWGLDNMICLLQGLAQS		

Further analysis of the NOV123a protein yielded the following properties shown in Table 123B.

465

Table 123B. Protein Sequence Properties NOV123a	
PSort analysis:	0.5663 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV123a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 123C.

Table 123C. Geneseq Results for NOV123a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV123a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41393	Human polypeptide SEQ ID NO 6324 - Homo sapiens, 687 aa. [WO200153312-A1, 26-JUL-2001]	1..683 5..687	682/683 (99%) 682/683 (99%)	0.0
AAM39607	Human polypeptide SEQ ID NO 2752 - Homo sapiens, 711 aa. [WO200153312-A1, 26-JUL-2001]	12..683 36..711	642/680 (94%) 643/680 (94%)	0.0
AAE04364	Human kinase (PKIN)-5 - Homo sapiens, 798 aa. [WO200146397-A2, 28-JUN-2001]	273..682 380..793	219/432 (50%) 285/432 (65%)	e-117

In a BLAST search of public sequence databases, the NOV123a protein was found to have homology to the proteins shown in the BLASTP data in Table 123D.

Table 123D. Public BLASTP Results for NOV123a				
Protein Accession Number	Protein/Organism/Length	NOV123a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96DU7	INOSITOL 1,4,5-TRISPHOSPHATE 3-KINASE C - Homo sapiens (Human), 683 aa.	1..683 1..683	683/683 (100%) 683/683 (100%)	0.0

466

Q9Y475	INOSITOL 1,4,5-TRISPHOSPHATE 3-KINASE ISOENZYME (EC 2.7.1.127) - Homo sapiens (Human), 604 aa (fragment).	83..683 4..604	601/601 (100%) 601/601 (100%)	0.0
S17682	1D-myo-inositol-trisphosphate 3-kinase (EC 2.7.1.127) B - human, 472 aa.	273..682 54..467	219/432 (50%) 285/432 (65%)	e-117
CAB65055	INOSITOL 1,4,5-TRISPHOSPHATE 3-KINASE B - Homo sapiens (Human), 946 aa.	273..682 528..941	219/432 (50%) 285/432 (65%)	e-117
Q96JS1	INOSITOL 1,4,5-TRISPHOSPHATE 3-KINASE, ISOFORM B (EC 2.7.1.127) - Homo sapiens (Human), 946 aa.	273..682 528..941	219/432 (50%) 285/432 (65%)	e-117

PFam analysis predicts that the NOV123a protein contains the domains shown in the Table 123E.

Table 123E. Domain Analysis of NOV123a			
Pfam Domain	NOV123a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

Example 124.

The NOV124 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 124A.

Table 124A. NOV124 Sequence Analysis	
	SEQ ID NO: 343 1395 bp
NOV124a, CG59993-01 DNA Sequence	GGTAAGACGACCTCTGGATGCTCACCCTGCCCTCTTCACCTCTCGTCCCCAGCTGTTT CCTCTGCCACCATGGAGGAACATTTTCAAGAGGAACCAGGAGCCTATTGTGGCTCCTGC CACCACCACCGCCACGATGCCCATTGGACCCGTGGACAACTCCACTGAGAGTGGGGGT GCTGGGGAGAGCCAGGAGGACATGTTTGCCAAACTGAAGGAGAAGTTATTCAATGAGA TAAACAAGATTCCCTTACCACCCTGGGCACTGATCGCCATTGCTGTGGTTGCTGGGCT CCTGCTTCTCACCTGCTGCTTCTGCATCTGCAAGAAATGCTGCTGCAAGAAGAAGAAG AACAAGAAGGAGAAGGGCAAAGGCATGAAGAATGCCATGAACATGAAGGACATGAAAG GGGGTCAGGATGACGACGACGCAGAGACAGGCCTGACTGAGGGGGAAGGTGAAGGGGA GGAGGAGAAAGAGCCAGAGAACCTGGGCAAACTGCAGTTTTCCCTGGACTATGATTTT CAGGCTAATCAGCTTACTGTGGGCGTTCTGCAGGCTGCTGAACTGCCTGCCCTGGACA TGGGGAGGCACCTCAGACCCTTATGTCAAGGTCTTCCTCCTTCCTGACAAGAAGAAGAA ATATGAGACCAAAGTCCATCGGAAGACACTGAACCCTGCCTTCAATGAAACCTTCACC TTCAAGGTGCCATACCAGGAGCTTGGGGGCAAAACTCTGGTGATGGCCATCTATGACT TTGACCGCTTCTCCAAACATGACATCATTGGAGAGGTAAAGGTGCCTATGAACACAGT GGACCTCGGCCAGCCCATTGAGGAGTGGAGAGACCTGCAAGGCGGGGAAAAGGAGGAG CCGGAGAAGCTGGGCGACATCTGCACCTCCCTGCGCTATGTGCCCACGGCCGGGAAGC TCACTGTCTGCATCCTGGAGGCTAAGAACCTCAAGAAGATGGACGTGGGCGGCCTTTC AGACCCGTACGTGAAGATCCACCTGATGCAGAATGGCAAGAGGCTCAAGAAGAAGAAG ACAACCGTGAAGAAGAAGACCCTGAACCCATACTTCAACGAGTCCTTCAGCTTTGAGA

467

	TCCCCTTCGAGCAGATTCAGAAAGTCCAGGTAGTGGTCACCGTGCTGGACTATGACAA GCTGGGCAAGAACGAAGCCATAGGCAAGATCTTCGTGGGCAGCAATGCCACGGGCACA GAGCTGCGGCACTGGTCCGACATGCTGGCCAACCCCCGGAGGCCCATCGCCCAGTGGC ACTCGCTCAAGCCTGAGGAGGAGGTGGATGCACTCCTGGGCAAGAACAAG<u>TAG</u>ACAGC <u>AGCGGCTGGGACCCCACACCTTTCACGGACACTGACAAGATCCAGAGCTATCAATACC TCA</u>
	ORF Start: ATG at 70	ORF Stop: TAG at 1327
	SEQ ID NO: 344	419 aa	MW at 46871.8kD
NOV124a, CG59993-01 Protein Sequence	MRNIFKRNQEPIVAPATTTATMPIGPVDNSTESGGAGESQEDMFAKLKEKLFNEINKI PLPPWALIAIAVVAGLLLLTCCFCICKKCCCKKKKNKKEKGKGMKNAMNMKDMKGGQD DDDAETGLTEGEGEGEEEKEPENLGKLQFSLDYDFQANQLTVGVLQAAELPALDMGGT SDPYVKVFLLPDKKKKYETKVHRKTLNPAFNETFTFKVPYQELGGKTLVMAIYDFDRF SKHDIIGEVKVPMNTVDLGQPIEEWRDLQGGEKEEPEKLGDICTSLRYVPTAGKLTVC ILEAKNLKKMDVGGLSDPYVKIHLMQNGKRLKKKKTTVKKKTLNPYFNESFSFEIPFE QIQKVQVVVTVLDYDKLGKNEAIGKIFVGSNATGTELRHWSDMLANPRRPIAQWHSLK PEEEVDALLGKNK
	SEQ ID NO: 345	1338 bp
NOV124b, CG59993-02 DNA Sequence	<u>CCACC</u>ATGAGGAACATTTTCAAGAGGAACCAGGAGCCTATTGTGGCTCCTGCCACCAC CACCGCCACGATGCCCATTGGACCCGTGGACAACTCCACTGAGAGTGGGGGTGCTGGG GAGAGTCAGGAGGACATGTTTGCCAAACTGAAGGAGAAGTTATTCAATGAGATAAACA AGATTCCCTTACCACCCTGGGCACTGATCGCCATTGCTGTGGTTGCTGGGCTCCTGCT TCTCACCTGCTGCTTCTGCATCTGCAAGAAATGCTGCTGCAAGAAGAAGAAGAACAAG AAGGAGAAGGGCAAAGGTATGAAGAATGCCATGAACATGAAGGACATGAAAGGGGGTC AGGATGACGACGACGCAGAGACAGGCCTGACTGAGGGGGAAGGTGAAGGGGAGGAGGA GAAAGAGCCAGAGAACCTGGGCAAACTGCAGTTTTCCCTGGACTATGATTTTCAGGCT AATCAGCTTACTGTGGGCGTTCTGCAGGCTGCTGAACTGCCTGCCCTGGACATGGGAG GCACCTCAGACCCTTATGTCAAGGTCTTCCTCCTTCCTGACAAGAAGAAGAAATATGA GACCAAAGTCCATCGGAAGACACTGAACCCTGCCTTCAATGAAACCTTCACCTTCAAG GTGCCATACCAGGAGCTTGGGGGCAAAACTCTGGTGATGGCCATCTATGACTTTGACC GCTTCTCCAAACATGACATCATTGGAGAGGTAAAGGTGCCTATGAACACAGTGGACCT CGGCCAGCCCATTGAGGAGTGGAGAGACCTGCAAGGCGGGGAAAAGGAGGAGCCGGAG AAGCTGGGCGACATCTGCACCTCCCTGCGCTATGTGCCCACGGCCGGGAAGCTCACTG TCTGCATCCTGGAGGCTAAGAACCTCAAGAAGATGGACGTGGGCGGCCTTTCAGACCC GTACGTGAAGATCCACCTGATGCAGAATGGCAAGAGGCTCAAGAAGAAGAAGACAACC ATGAAGAAGAAGACCCTGAACCCATACTTCAACGAGTCCTTCAGCTTTGAGATCCCCT TCGAGCAGATTCAGAAAGTCCAGGTAGTGGTCACCGTGCTGGACTATGACAAGCTGGG CAAGAACGAAGCCATAGGCAAGATCTTCGTGGGCAGCAATGCCACGGGCACACTCGC CGGCACTGGTCCGACATGCTGGCCAACCCCCGGAGGCCCATCGCCCAGTGGCACTCGC TCAAGCCTGAGGAGGAGGTGGGTGCACTCCTGGGCAAGAACAAGTAGACAGCAGCGGC <u>TGGGACCCCACACCTTTCACGGACACTGACAAGATCCAGAGCTATCAATAAGGTGTAG GCGG</u>
	ORF Start: ATG at 6	ORF Stop: TAG at 1263
	SEQ ID NO: 346	419 aa	MW at 46845.9kD
NOV124b, CG59993-02 Protein Sequence	MRNIFKRNQEPIVAPATTTATMPIGPVDNSTESGGAGESQEDMFAKLKEKLFNEINKI PLPPWALIAIAVVAGLLLLTCCFCICKKCCCKKKKNKKEKGKGMKNAMNMKDMKGGQD DDDAETGLTEGEGEGEEEKEPENLGKLQFSLDYDFQANQLTVGVLQAAELPALDMGGT SDPYVKVFLLPDKKKKYETKVHRKTLNPAFNETFTFKVPYQELGGKTLVMAIYDFDRF SKHDIIGEVKVPMNTVDLGQPIEEWRDLQGGEKEEPEKLGDICTSLRYVPTAGKLTVC ILEAKNLKKMDVGGLSDPYVKIHLMQNGKRLKKKKTTMKKKTLNPYFNESFSFEIPFE QIQKVQVVVTVLDYDKLGKNEAIGKIFVGSNATGTELRHWSDMLANPRRPIAQWHSLK PEEEVGALLGKNK

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 124B.

Table 124B. Comparison of NOV124a against NOV124b.		
Protein Sequence	NOV124a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV124b	1..419 1..419	335/419 (79%) 335/419 (79%)

468

Further analysis of the NOV124a protein yielded the following properties shown in Table 124C.

Table 124C. Protein Sequence Properties NOV124a	
PSort analysis:	0.8202 probability located in mitochondrial inner membrane; 0.6000 probability located in endoplasmic reticulum (membrane); 0.3500 probability located in nucleus; 0.3034 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV124a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 124D.

Table 124D. Geneseq Results for NOV124a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV124a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR97722	Mouse inositol polyphosphate binding protein IP4-BP - Mus musculus, 422 aa. [JP08092290-A, 09-APR-1996]	1..419 1..422	412/422 (97%) 414/422 (97%)	0.0
AAU19715	Human novel extracellular matrix protein, Seq ID No 365 - Homo sapiens, 461 aa. [WO200155368-A1, 02-AUG-2001]	128..405 169..447	141/280 (50%) 201/280 (71%)	2e-80
AAU19714	Human novel extracellular matrix protein, Seq ID No 364 - Homo sapiens, 295 aa. [WO200155368-A1, 02-AUG-2001]	141..409 11..281	140/273 (51%) 193/273 (70%)	3e-74
AAW87702	A human membrane fusion protein designated SYTAX2 - Homo sapiens, 375 aa. [WO9856813-A2, 17-DEC-1998]	59..407 31..364	146/352 (41%) 220/352 (62%)	4e-73
AAO05534	Human polypeptide SEQ ID NO 19426 - Homo sapiens, 149 aa. [WO200164835-A2, 07-SEP-2001]	33..164 15..149	127/135 (94%) 131/135 (96%)	5e-70

In a BLAST search of public sequence databases, the NOV124a protein was found to have homology to the proteins shown in the BLASTP data in Table 124E.

Table 124E. Public BLASTP Results for NOV124a				
Protein Accession Number	Protein/Organism/Length	NOV124a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P29101	Synaptotagmin II (SytII) - Rattus norvegicus (Rat), 422 aa.	1..419 1..422	411/422 (97%) 414/422 (97%)	0.0
A55417	synaptotagmin II - mouse, 422 aa.	1..419 1..422	412/422 (97%) 414/422 (97%)	0.0
P46097	Synaptotagmin II (SytII) - Mus musculus (Mouse), 422 aa.	1..419 1..422	411/422 (97%) 413/422 (97%)	0.0
P24506	Synaptotagmin B (Synaptic vesicle protein O-P65-B) - Discopyge ommata (Electric ray), 439 aa.	10..419 27..439	341/413 (82%) 366/413 (88%)	0.0
P46096	Synaptotagmin I (SytI) (p65) - Mus musculus (Mouse), 421 aa.	10..419 8..421	323/418 (77%) 353/418 (84%)	0.0

PFam analysis predicts that the NOV124a protein contains the domains shown in the Table 124F.

Table 124F. Domain Analysis of NOV124a			
Pfam Domain	NOV124a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Adeno_E3_CR2: domain 1 of 1	62..108	16/50 (32%) 26/50 (52%)	6.5
C2: domain 1 of 2	156..242	54/97 (56%) 81/97 (84%)	1.8e-42
C2: domain 2 of 2	287..375	44/97 (45%) 80/97 (82%)	2.9e-39

Example 125.

The NOV125 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 125A.

Table 125A. NOV125 Sequence Analysis

	SEQ ID NO: 347	3226 bp

NOV125a, CG59991-01 DNA Sequence	GGACCACTTCTGATGCATCTCTGGGTCCCAACACTATCCACTGCAAGGCCTCGAAACA GGGGGGCCAG**ATG**GGGACCCCCATTTAGCACAAGAGAGACGTCCACACTCTGTGAGCCC AAAGGGAGAAGGCTCAGGCCACGGCAGAGACGGAACCAGGAAAACGTCACGAAAAACA GCCTCAAGTTGCCAGGTCCCTTGCAGGAACAGACAGGCCTGGGGCCGCCCCACCTGGG CTCAGAGCTTGGGCTGCATGGAGGTGACACATGGGACTACAAGAGTCACGTGATGACC AAATTCGCTGAGGAGGAGGATGTACGTCGTAGTTTTGAAAACACTGCTGCTGACTGGC CGGAAATGCAAACGTTGGCTGGTGCTTTTGATTCAGACCGGTGGGGCTTCCGGCCTCG CACGGTGGTTCTGCACGGAAAGTCAGGAATTGGGAAATCGGCTCTAGCCAGAAGGATC GTGCTGTGCTGGGCGCAAGGTGGACTCTACCAGGGAATGTTCTCCTACGTCTTCTTCC TCCCCGTTAGAGAGATGCAGCGGAAGAAGGAGAGCAGTGTCACAGAGTTCATCTCCAG GGAGTGGCCAGACTCCCAGGCTCCGGTGACGGAGATCATGTCCCGACCAGAAAGGCTG TTGTTCATCATTGACGGTTTCGATGACCTGGGCTCTGTCCTCAACAATGACACAAAGC TCTGCAAAGACTGGGCTGAGAAGCAGCCTCCGTTCACCCTCATACGCAGTCTGCTGAG GAAGGTCCTGCTCCCTGAGTCCTTCCTGATCGTCACCGTCAGAGACGTGGGCACAGAG AAGCTCAAGTCAGAGGTCGTGTCTCCCCGTTACCTGTTAGTTAGAGGAATCTCCGGGG AACAAAGAATCCACTTGCTCCTTGACGCGGGATTGGTGAGCATCAGAAGACACAAGG GTTGCGTGCGATCATGAACAACCGTGAGCTGCTCGACCAGTGCCAGGTGCCCGCCGTG GGCTCTCTCATCTGCGTGGCCCTGCAGCTGCAGGACGTGGTGGGGGAGAGCGTCGCCC CCTTCAACCAAACGCTCACAGGCCTGCACGCCGCTTTTGTGTTTCATCAGCTCACCCC TCGAGGCGTGGTCCGGCGCTGTCTCAATCTGGAGGAAAGAGTTGTCCTGAAGCGCTTC TGCCGTATGGCTGTGGAGGGAGTGTGGAATAGGAAGTCAGTGTTTGACGGTGACGACC TCATGGTTCAAGGACTCGGGGAGTCTGAGCTCCGTGCTCTGTTTCACATGAACATCCT TCTCCCAGACAGCCACTGTGAGGAGTACTACACCTTCTTCCACCTCAGTCTCCAGGAC TTCTGTGCCGCCTTGTACTACGTGTTAGAGGGCCTGGAAATCGAGCCAGCTCTCTGCC CTCTGTACGTTGAGAAGACAAAGAGGTCCATGGAGCTTAAACAGGCAGGCTTCCATAT CCACTCGCTTTGGATGAAGCGTTTCTTGTTTGGCCTCGTGAGCGAAGACGTAAGGAGG CCACTGGAGGTCCTGCTGGGCTGTCCCGTTCCCCTGGGGGTGAAGCAGAAGCTTCTGC ACTGGGTCTCTCTGTTGGGTCAGCAGCCTAATGCCACCACCCCAGGAGACACCCTGGA CGCCTTCCACTGTCTTTTCGAGACTCAAGACAAAGAGTTTGTTCGCTTGGCATTAAAC AGCTTCCAAGAAGTGTGGCTTCCGATTAACCAGAACCTGGACTTGATAGCATCTTCCT TCTGCCTCCAGCACTGTCCGTATTTGCGGAAAATTCGGGTGGATGTCAAAGGGATCTT CCCAAGAGATGAGTCCGCTGAGGCATGTCCTGTGGTCCCTCTATGGATGCGGGATAAG ACCCTCATTGAGGAGCAGTGGGAAGATTTCTGCTCCATGCTTGGCACCCACCCACACC TGCGGCAGCTGGACCTGGGCAGCAGCATCCTGACAGAGCGGGCCATGAAGACCCTGTG TGCCAAGCTGAGGCATCCCACCTGCAAGATACAGACCCTGATGTTTAGAAATGCACAG ATTACCCCTGGTGTGCAGCACCTCTGGAGAATCGTCATGGCCAACCGTAACCTAAGAT CCCTCAACTTGGGAGGCACCCACCTGAAGGAAGAGGATGTAAGGATGGCGTGTGAAGC CTTAAAACACCCAAAATGTTTGTTGGAGTCTTTGAGGCTGGATTGCTGTGGATTGACC CATGCCTGTTACCTGAAGATCTCCCAAATCCTTACGACCTCCCCCAGCCTGAAATCTC TGAGCCTGGCAGGAAACAAGGTGACAGACCAGGGAGTAATGCCTCTCAGTGATGCCTT GAGAGTCTCCCAGTGCGCCCTGCAGAAGCTGATACTGGAGGACTGTGGCATCACAGCC ACGGGTTGCCAGAGTCTGGCCTCAGCCCTCGTCAGCAACCGGAGCTTGACACACCTGT GCCTATCCAACAACAGCCTGGGGAACGAAGGTGTAAATCTACTGTGTCGATCCATGAG GCTTCCCCACTGTAGTCTGCAGAGGCTGATGCTGAATCAGTGCCACCTGGACACGGCT GGCTGTGGTTTTCTTGCACTTGCGCTTATGGGTAACTCATGGCTGACGCACCTGAGCC TTAGCATGAACCCTGTGGAAGACAATGGCGTGAAGCTTCTGTGCGAGGTCATGAGAGA ACCATCTTGTCATCTCCAGGACCTGGAGTTGGTAAAGTGTCATCTCACCGCCGCGTGC TGTGAGAGTCTGTCCTGTGTGATCTCGAGGAGCAGACACCTGAAGAGCCTGGATCTCA CGGACAATGCCCTGGGTGACGGTGGGGTTGCTGCACTGTGCGAGGGACTGAAGCAAAA GAACAGTGTTCTGACGAGACTCGGGTTGAAGGCATGTGGACTGACTTCTGATTGCTGT GAGGCACTCTCCTTGGCCCTTTCCTGCAACCGGCATCTGACCAGTCTAAACCTGGTGC AGAATAACTTCAGTCCCAAAGGAATGATGAAGCTGTGTTCGGCCTTTGCCTGTCCCAC GTCTAACTTACAGATAATTGGGCTGTGGAAATGGCAGTACCCTGTGCAAATAAGGAAG CTGCTGGAGGAAGTGCAGCTACTCAAGCCCCGAGTCGTAATTGACGGTAGTTGGCATT CTTTTGATGAAGATGACCGGTACTGGTGGAAAAAC**TGA**AGATACGGAAACCTGCCCCA CTCACACCCATCTGATGGAGGAACTTTAAACGCTGT

	ORF Start: ATG at 69	ORF Stop: TGA at 3168	
	SEQ ID NO: 348	1033 aa	MW at 116310.7kD

NOV125a, CG59991-01 Protein Sequence	MGPPFSTRETSTLCEPKGRRLRPRQRRNQENVTKNSLKLPGPLQEQTGLGPPHLGSEL GLHGGDTWDYKSHVMTKFAEEEDVRRSFENTAADWPEMQTLAGAFDSDRWGFRPRTVV LHGKSGIGKSALARRIVLCWAQGGLYQGMFSYVFFLPVREMQRKKESSVTEFISREWP DSQAPVTEIMSRPERLLFIIDGFDDLGSVLNNDTKLCKDWAEKQPPFTLIRSLLRKVL LPESFLIVTVRDVGTEKLKSEVVSPRYLLVRGISGEQRIHLLLERGIGEHQKTQGLRA IMNNRELLDQCQVPAVGSLICVALQLQDVVGESVAPFNQTLTGLHAAFVFHQLTPRGV VRRCLNLEERVVLKRFCRMAVEGVWNRKSVFDGDDLMVQGLGESELRALFHMNILLPD SHCEEYYTFFHLSLQDFCAALYYVLEGLEIEPALCPLYVEKTKRSMELKQAGFHIHSL WMKRFLFGLVSEDVRRPLEVLLGCPVPLGVKQKLLHWVSLLGQQPNATTPGDTLDAFH CLFETQDKEFVRLALNSFQEVWLPINQNLDLIASSFCLQHCPYLRKIRVDVKGIFPRD ESAEACPVVPLWMRDKTLIEEQWEDFCSMLGTHPHLRQLDLGSSILTERAMKTLCAKL

471

| | RHPTCKIQTLMFRNAQITPGVQHLWRIVMANRNLRSLNLGGTHLKEEDVRMACEALKH PKCLLESLRLDCCGLTHACYLKISQILTTSPSLKSLSLAGNKVTDQGVMPLSDALRVS QCALQKLILEDCGITATGCQSLASALVSNRSLTHLCLSNNSLGNEGVNLLCRSMRLPH CSLQRLMLNQCHLDTAGCGFLALALMGNSWLTHLSLSMNPVEDNGVKLLCEVMREPSC HLQDLELVKCHLTAACCESLSCVISRSRHLKSLDLTDNALGDGGVAALCEGLKQKNSV LTRLGLKACGLTSDCCEALSLALSCNRHLTSLNLVQNNFSPKGMMKLCSAFACPTSNL QIIGLWKWQYPVQIRKLLEEVQLLKPRVVIDGSWHSFDEDDRYWWKN |

Further analysis of the NOV125a protein yielded the following properties shown in Table 125B.

Table 125B. Protein Sequence Properties NOV125a	
PSort analysis:	0.7600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV125a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 125C.

Table 125C. Geneseq Results for NOV125a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV125a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE07514	Human PYRIN-1 protein - Homo sapiens, 1034 aa. [WO200161005-A2, 23-AUG-2001]	103..934 207..1003	276/843 (32%) 445/843 (52%)	e-126
AAE07513	Human nucleotide binding site 1 (NBS-1) protein - Homo sapiens, 1033 aa. [WO200161005-A2, 23-AUG-2001]	114..935 180..990	281/839 (33%) 431/839 (50%)	e-120
AAU07878	Polypeptide sequence for mammalian Spg65 - Mammalia, 748 aa. [WO200166752-A2, 13-SEP-2001]	207..963 9..748	218/766 (28%) 380/766 (49%)	7e-95
AAE06758	Human G-protein coupled receptor-8 (GCREC-8) protein - Homo sapiens, 1473 aa. [WO200157085-A2, 09-AUG-2001]	21..764 219..959	235/772 (30%) 380/772 (48%)	3e-88
AAB62571	Human CARD-7 polypeptide -	21..764 219..959	235/772 (30%) 380/772 (48%)	3e-88

472

	[WO200130813-A1, 03-MAY-2001]			

In a BLAST search of public sequence databases, the NOV125a protein was found to have homology to the proteins shown in the BLASTP data in Table 125D.

Table 125D. Public BLASTP Results for NOV125a				
Protein Accession Number	Protein/Organism/Length	NOV125a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9JLR2	MATERNAL-ANTIGEN-THAT-EMBRYOS-REQUIRE PROTEIN - Mus musculus (Mouse), 1111 aa.	24..1033 104..1111	548/1019 (53%) 716/1019 (69%)	0.0
Q9R1M5	MATER PROTEIN - Mus musculus (Mouse), 1111 aa.	24..1033 104..1111	547/1019 (53%) 716/1019 (69%)	0.0
AAL35293	NALP4 - Homo sapiens (Human), 994 aa.	63..958 94..981	291/907 (32%) 473/907 (52%)	e-133
Q96MN2	CDNA FLJ32126 FIS, CLONE PEBLM2000112, WEAKLY SIMILAR TO HOMO SAPIENS NUCLEOTIDE-BINDING SITE PROTEIN 1 MRNA - Homo sapiens (Human), 919 aa.	63..958 19..906	291/907 (32%) 473/907 (52%)	e-133
AAL12497	CRYOPYRIN - Homo sapiens (Human), 1034 aa.	103..934 207..1003	276/843 (32%) 445/843 (52%)	e-125

PFam analysis predicts that the NOV125a protein contains the domains shown in the Table 125E.

Table 125E. Domain Analysis of NOV125a			
Pfam Domain	NOV125a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRR: domain 1 of 6	671..695	6/25 (24%) 16/25 (64%)	1.6e+02
LRR: domain 2 of 6	728..752	7/27 (26%) 17/27 (63%)	2.3e+02

473

LRR: domain 3 of 6	785..809	7/26 (27%) 19/26 (73%)	1.6e+02
LRR: domain 4 of 6	814..836	6/25 (24%) 14/25 (56%)	4.3e+02
LRR: domain 5 of 6	899..923	8/26 (31%) 20/26 (77%)	27
LRR: domain 6 of 6	956..977	7/25 (28%) 16/25 (64%)	2.9e+02

Example 126.

The NOV126 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 126A.

Table 126A. NOV126 Sequence Analysis		
	SEQ ID NO: 349	2310 bp
NOV126a, CG59987-01 DNA Sequence	CCGCGCCTCAGTCCGCCGTCCGCCCTCCGCGCCCGCGCCGCTAGC**ATG**ACCGACGCGC TGTTGCCCGCGGCCCCCCAGCCGCTGGAGAAGGAGAACGACGGCTACTTTCGGAAGGG CTGTAATCCCCTTGCACAAACCGGCCGGAGTAAATTGCAGAATCAAAGAGCTGCTTTG AATCAGCAGATCCTGAAAGCCGTGCGGATGAGGACCGGAGCCGGAAAACCTTCTGAAAG TGGCCACAAACTCAAAGGTGCGGGAGCAAGTCGGCTGGAGCTGAGCTTCGTCAACTC AGACCTGCAGATGCTCAAGGAAGAGCTGGAGGGGCTGAACATCTCGGTGGGCGTCTAT CAGAACACAGAGGAGGCATTTACGATTCCCCTGATTCCTCTTGGCCTGAAGGAAACGA AAGACGTCGACTTTGCAGTCGTCCTCAAGGATTTTATCCTGGAACATTACAGTGAAGA TGGCTATTTATATGAAGATGAAATTGCAGATCTTATGGATCTGAGACAAGCTTGTCGG ACGCCTAGCCGGGATGAGGCCGGGGTGGAACTGCTGATGACATACTTCATCCAGCTGG GCTTTGTCGAGAGTCGATTCTTCCCGCCCACACGGCAGATGGGACTCCTGTTCACCTG GTATGACTCTCTCACCGGGGTTCCGGTCAGCCAGCAGAACCTGCTGCTGGAGAAGGCC AGTGTCCTGTTCAACACTGGGGCCCTCTACACCCAGATTGGGACCCGGTGTGATCGGC AGACGCAGGCTGGGCTGGAGAGTGCCATAGATGCCTTTCAGAGAGCCGCAGGGGTTT AAAATTACCTGAAAGACACATTTACCCATACTCCAAGTTACGACATGAGCCCTGCCATG CTCAGCGTGCTCGTCAAAATGATGCTTGCACAAGCCCAAGAAAGCGTGTTTGAGAAAA TCAGCCTTCCTGGGATCCGGAATGAATTCTTCATGCTGGTGAAGGTGGCTCAGGAGGC TGCTAAGGTGGGAGAGGTCTACCAACAGCTACACGCAGCCATGAGCCAGGCGCCGGTG AAAGAGAACATCCCCTACTCCTGGGCCAGCTTAGCCTGCGTGAAGGCCCACCACTACG CGGCCCTGGCCCACTACTTCACTGCCATCCTCCTCATCGACCACCAGGTGAAGCCAGG CACGGATCTGGACCACCAGGAGAAGTGCCTGTCCCAGCTCTACGACCACATGCCAGAG GGGCTGACACCCTTGGCCACACTGAAGAATGATCAGCAGCGCCGACAGCTGGGGAAGT CCCACTTGCGCAGAGCCATGGCTCATCACGAGGAGTCGGTGCGGGAGGCCAGCCTCTG CAAGAAGCTGCGGAGCATTGAGGTGCTACAGAAGGTGCTGTGTGCCGCACAGGAACGC TCCCGGCTCACGTACGCCCAGCACCAGGAGGAGGATGACCTGCTGAACCTGATCGACG CCCCCAGAGTGTTGTTGCTAAAACTGAGCAAGAGGTTGACATTATATTGCCCCAGTTC TCCAGCTGACAGTCACGGACTTCTTCCAGAAGCTGGGCCCTTATCTGTGCTGTCGGCT AACAAGCGGTGGACGCCTCCTCGAAGCATCCGCTTCACTGCAGAAGAAGGGGACTTGG GGTTCACCTTGAGAGGGAACGCCCCCGTTCAGGTTCACTTCCTGGATCCTTACTGCTC TGCCTCGGTGGCAGGAGCCCGGGAAGGAGATTATATTGTCTCCATTCAGCTTGTGGAT TGTAAGTGGCTGACGCTGAGTGAGGTTATGAAGCTGCTGAAGAGCTTTGGCGAGGACG AGATCGAGATGAAAGTCGTGAGCCTCCTGGACTCCACATCATCCATGCATAATAAGAG TGCCACATACTCCGTGGGAATGCAGAAAACGTACTCCATGATCTGCTTAGCCATTGAT GATGACGACAAAACTGATAAAACCAAGAAAATCTCCAAGAAGCTTTCCTTCCTGAGTT GGGGCACCAACAAGAACAGACAGAAGTCAGCCAGCACCTTGTGCCTCCCATCGGTCGG GGCTGCACGGCCTCAGGTCAAGAAGAAGCTGCCCTCCCCTTTCAGCCTTCTCAACTCA GACAGTTCTTGGTACTAA TGTGAGGAAACAAACATGTTCAGGCCCCGAACATTTCCGG TGCTGACTCGGCCTTAAACGTTTGTGCCATAATGGAAAATATCTATCTATCTGTTCTC AAATCCTGTTTTTCTCATAGTGTAAACTCACATTTGATGTGTTTTTATGAAGGAAAGT AACCAAGAAACCTCTAGGAATTAGTGAAAAAAGAACTTTTTTGAGGTG

	ORF Start: ATG at 46	ORF Stop: TAA at 2104	
	SEQ ID NO: 350	686 aa	MW at 76812.3kD
NOV126a, CG59987-01 Protein Sequence	MTDALLPAAPQPLEKENDGYFRKGCNPLAQTGRSKLQNQRAALNQQILKAVRMRTGAE NLLKVATNSKVREQVRLELSFVNSDLQMLKEELEGLNISVGVYQNTEEAFTIPLIPLG LKETKDVDFAVVLKDFILEHYSEDGYLYEDEIADLMDLRQACRTPSRDEAGVELLMTY		

474

	FIQLGFVESRFFPPTRQMGLLFTWYDSLTGVPVSQQNLLLEKASVLFNTGALYTQIGT RCDRQTQAGLESAIDAFQRAAGVLNYLKDTFTHTPSYDMSPAMLSVLVKMMLAQAQES VFEKISLPGIRNEFFMLVKVAQEAAKVGEVYQQLHAAMSQAPVKENIPYSWASLACVK AHHYAALAHYFTAILLIDHQVKPGTDLDHQEKCLSQLYDHMPEGLTPLATLKNDQQRR QLGKSHLRRAMAHHEESVREASLCKKLRSIEVLQKVLCAAQERSRLTYAQHQEEDDLL NLIDAPRVLLLKLSKRLTLYCPSSPADSHGLLPEAGPLSVLSANKRWTPPRSIRFTAE EGDLGFTLRGNAPVQVHFLDPYCSASVAGAREGDYIVSIQLVDCKWLTLSEVMKLLKS FGEDEIEMKVVSLLDSTSSMHNKSATYSVGMQKTYSMICLAIDDDDKTDKTKKISKKL SFLSWGTNKNRQKSASTLCLPSVGAARPQVKKKLPSPFSLLNSDSSWY
	SEQ ID NO: 351 2109 bp
NOV126b, CG59987-02 DNA Sequence	CGCCGCTAGC**ATGA**CCGACGCGCTGTTGCCCGCGGCCCCCCAGCCGCTGGAGAAGGAG AACGACGGCTACTTTCGGAAGGGCTGTAATCCCCTTGCACAAACCGGCCGGAGTAAAT TGCAGAATCAAAGAGCTGCTTTGAATCAGCAGATCCTGAAAGCCGTGCGGATGAGGAC CGGAGCGGAAAACCTTCTGAAAGTGGCCACAAACTCAAAGGTGCGGGAGCAAGTGCGG CTGGAGCTGAGCTTCGTCAACTCAGACCTGCAGATGCTCAAGGAAGAGCTGGAGGGGC TGAACATCTCGGTGGGCGTCTATCAGAACACAGAGGAGGCATTTACGATTCCCCTGAT TCCTCTTGGCCTGAAGGAAACGAAAGACGTCGACTTTGCAGTCGTCCTCAAGGATTTT ATCCTGGAACATTACAGTGAAGATGGCTATTTATATGAAGATGAAATTGCAGATCTTA TGGATCTGAGACAAGCTTGTCGGACGCCTAGCCGGGATGAGGCCGGGGTGGAACTGCT GATGACATACTTCATCCAGCTGGGCTTTGTCGAGAGTCGATTCTTCCCGCCCACACGG CAGATGGGACTCCTGTTCACCTGGTATGACTCTCTCACCGGGGTTCCGGTCAGCCAGC AGAACCTGCTGCTGGAGAAGGCCAGTGTCCTGTTCAACACTGGGGCCCTCTACACCCA GATTGGGACCCGGTGCGATCGGCAGACGCAGGCTGGGCTGGAGAGTGCCATAGATGCC TTTCAGAGAGCCGCAGGGGTTTTAAATTACCTGAAAGACACATTTACCCATACTCCAA GTTACGACATGAGCCCTGCCATGCTCAGCGTGCTCGTCAAAATGATGCTTGCACAAGC CCAAGAAAGCGTGTTTGAGAAAATCAGCCTTCCTGGGATCCGGAATGAATTCTTCATG CTGGTGAAGGTGGCTCAGGAGGCTGCTAAGGTGGGAGAGGTCTACCAACAGCTACACG CAGCCATGAGCCAGGCGCCGGTGAAAGAGAACATCCCCTACTCCTGGGCCAGCTTAGC CTGCGTGAAGGCCCACCACTACGCGGCCCTGGCCCACTACTTCACTGCCATCCTCCTC ATCGACCACCAGGTGAAGCCAGGCACGGATCTGGACCACCAGGAGAAGTGCCTGTCCC AGCTCTACGACCACATGCCAGAGGGGCTGACACCCTTGGCCACACTGAAGAATGATCA GCAGCGCCGACAGCTGGGGAAGTCCCACTTGCGCAGAGCCATGGCTCATCACGAGGAG TCGGTGCGGGAGGCAAGCCTCTGCAAGAAGCTGCGGAGCATTGAGGTGCTACAGAAGG TGCTGTGTGCCGCACAGGAACGCTCCCGGCTCACGTACGCCCAGCACCAGGAGGAGGA TGACCTGCTGAACCTGATCGACGCCCCCAGTGTTGTTGCTAAAACTGAGCAAGAGGTT GACATTATATTGCCCCAGTTCTCCAAGCTGACAGTCACGGACTTCTTCCAGAAGCTGG GCCCCTTATCTGTGTTTTCGGCTAACAAGCGGTGGACGCCTCCTCGAAGCATCCGCTT CACTGCAGAAGAAGGGGACTTGGGGTTCACCTTGAGAGGGAACGCCCCCGTTCAGGTT CACTTCCTGGATCCTTACTGCTCTGCCTCGGTGGCAGGAGCCCGGGAAGGAGATTATA TTGTCTCCATTCAGCTTGTGGATTGTAAGTGGCTGACGCTGAGTGAGGTTATGAAGCT GCTGAAGAGCTTTGGCGAGGACGAGATCGAGATGAAAGTCGTGAGCCTCCTGGACTCC ACATCATCCATGCATAATAAGAGTGCCACATACTCCGTGGGAATG**TAGA**AAACGTACT <u>CCATGATCTGCTTAGCCATTGATGATGACGACAAAACTGATAAAACCAAGAAAATCTC CAAGAAGCTTTCCTTCCTGAGTTGGGGCACCAACAAGAACAGACAGAAGTCAGCCAGC ACCTTGTGCCTCCCATCGGTCGGGGCTGCACGGCCTCAGGTCAAGAAGAAGCTGCCCT CCCCTTTCAGCCTTCTCAACTCAGACAGTTCTTGGTACTAATGTGAGGAAACAAACAT GTTCAGGCCCCGAACATTTCC</u>
	ORF Start: ATG at 11 **ORF Stop: TAG at 1844**
	SEQ ID NO: 352 611 aa MW at 68613.9kD
NOV126b, CG59987-02 Protein Sequence	MTDALLPAAPQPLEKENDGYFRKGCNPLAQTGRSKLQNQRAALNQQILKAVRMRTGAE NLLKVATNSKVREQVRLELSFVNSDLQMLKEELEGLNISVGVYQNTEEAFTIPLIPLG LKETKDVDFAVVLKDFILEHYSEDGYLYEDEIADLMDLRQACRTPSRDEAGVELLMTY FIQLGFVESRFFPPTRQMGLLFTWYDSLTGVPVSQQNLLLEKASVLFNTGALYTQIGT RCDRQTQAGLESAIDAFQRAAGVLNYLKDTFTHTPSYDMSPAMLSVLVKMMLAQAQES VFEKISLPGIRNEFFMLVKVAQEAAKVGEVYQQLHAAMSQAPVKENIPYSWASLACVK AHHYAALAHYFTAILLIDHQVKPGTDLDHQEKCLSQLYDHMPEGLTPLATLKNDQQRR QLGKSHLRRAMAHHEESVREASLCKKLRSIEVLQKVLCAAQERSRLTYAQHQEEDDLL NLIDAPSVVAKTEQEVDIILPQFSKLTVTDFFQKLGPLSVFSANKRWTPPRSIRFTAE EGDLGFTLRGNAPVQVHFLDPYCSASVAGAREGDYIVSIQLVDCKWLTLSEVMKLLKS FGEDEIEMKVVSLLDSTSSMHNKSATYSVGM

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 126B.

Table 126B. Comparison of NOV126a against NOV126b.		
Protein Sequence	**NOV126a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV126b	1..611 1..611	585/612 (95%) 590/612 (95%)

Further analysis of the NOV126a protein yielded the following properties shown in Table 126C.

Table 126C. Protein Sequence Properties NOV126a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV126a protein against the Geneseq database, a proprietary
database that contains sequences published in patents and patent publication, yielded
several homologous proteins shown in Table 126D.

Table 126D. Geneseq Results for NOV126a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV126a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU10192	Human prostate specific protein PSL22 - Homo sapiens, 686 aa. [WO200172962-A2, 04-OCT-2001]	1..686 1..686	660/687 (96%) 665/687 (96%)	0.0
AAB68561	Human GTP-binding associated protein #61 - Homo sapiens, 666 aa. [WO200105970-A2, 25-JAN-2001]	27..686 7..666	626/661 (94%) 633/661 (95%)	0.0
AAG64579	Human transcription termination factor binding protein 54 - Homo sapiens, 488 aa. [CN1297918-A, 06-JUN-2001]	201..686 3..488	458/487 (94%) 464/487 (95%)	0.0
AAB29661	Human histidine domain-protein tyrosine phosphatase, SEQ ID NO:2	110..357 7..253	82/252 (32%) 135/252 (53%)	3e-28

476

	[WO200063392-A1, 26-OCT-2000]			
AAU00869	Human cancer related protein 5 - Homo sapiens, 257 aa. [WO200118014-A1, 15-MAR-2001]	409..597 8..196	70/189 (37%) 102/189 (53%)	2e-27

In a BLAST search of public sequence databases, the NOV126a protein was found to have homology to the proteins shown in the BLASTP data in Table 126E.

Table 126E. Public BLASTP Results for NOV126a				
Protein Accession Number	Protein/Organism/Length	NOV126a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96RU1	RHOPHILIN-LIKE PROTEIN - Homo sapiens (Human), 685 aa.	1..686 1..685	627/688 (91%) 640/688 (92%)	0.0
Q9DBN2	1300002E07RIK PROTEIN - Mus musculus (Mouse), 686 aa.	1..686 1..686	573/687 (83%) 616/687 (89%)	0.0
Q61085	GTP-RHO binding protein 1 (Rhophilin) - Mus musculus (Mouse), 643 aa.	16..596 20..580	273/583 (46%) 361/583 (61%)	e-135
Q9XYY9	RHOPHILIN - Drosophila melanogaster (Fruit fly), 718 aa.	21..615 31..674	248/654 (37%) 363/654 (54%)	e-110
Q96PV9	KIAA1929 PROTEIN - Homo sapiens (Human), 410 aa (fragment).	23..366 17..362	178/346 (51%) 241/346 (69%)	1e-93

PFam analysis predicts that the NOV126a protein contains the domains shown in
5   the Table 126F.

Table 126F. Domain Analysis of NOV126a			
Pfam Domain	NOV126a Match Region	Identities/ Similarities for the Matched Region	Expect Value
HR1: domain 1 of 1	38..110	19/87 (22%) 53/87 (61%)	1.2e-05
BRO1: domain 1 of 1	111..263	60/172 (35%) 125/172 (73%)	3.8e-56

477

PDZ: domain 1 of 1	516..593	20/84 (24%) 53/84 (63%)	0.46

Example 127.

The NOV127 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 127A.

<table>
<tr><td colspan="3" align="center">Table 127A. NOV127 Sequence Analysis</td></tr>
<tr><td></td><td>SEQ ID NO: 353</td><td>3351 bp</td></tr>
<tr><td>NOV127a,
CG59971-01 DNA Sequence</td><td colspan="2">CGTCCCGTGGCCATGACGACCGCTCAGAGGGACTCCCTGTTGTGGAAGCTCGCGGGGT
TGCTGCGGGAGTCCGGTGATGTGGTCCTGTCTGGCTGTAGCACCCTGAGCCTGCTGAC
TCCCACACTGCAACAGCTGAACCACGTATTTGAGCTGCACCTGGGGCCATGGGGCCCT
GGCCAGACAGGCTTTGTGGCTCTGCCCTCCCATCCTGCCGACTCCCCTGTTATTCTTC
AGCTTCAGTTTCTCTTCGATGTGCTGCAGAAAACACTTTCACTCAAGCTGGTCCATGT
TGCTGGTCCTGGCCCCACAGGGCCCATCAAGATTTTCCCCTTCAAATCCCTTCGGCAC
CTGGAGCTCCGAGGTGTTCCCCTCCACTGTCTGCATGGCCTCCGAGGCATCTACTCCC
AGCTGGAGACCCTGATTTGCAGCAGGAGCCTCCAGGCATTAGAGGAGCTCCTCTCAGC
CTGCGGCGGCGACTTCTGCTCTGCCCTCCCTTGGCTGGCTCTGCTTTCTGCCAACTTC
AGCTACAATGCACTGACCGCCTTAGACAGCTCCCTGCGCCTCTTGTCAGCTCTGCGTT
TCTTGAACCTAAGCCACAATCAAGTCCAGGACTGTCAGGGATTCCTGATGGATTTGTG
TGAGCTCCACCATCTGGACATCTCCTATAATCGCCTGCATTTGGTGCCAAGAATGGGA
CCCTCAGGGGCTGCTCTGGGGGTCCTGATACTGCGAGGCAATGAGCTTCGGAGCCTGC
CAGGCCTAGAGCAGCTGAGGAATCTGCGGCACCTGGATTTGGCATACAACCTGCTGGA
AGGACACCGGGAGCTGTCACCACTGTGGCTGCTGGCTGAGCTCCGCAAGCTCTACCTG
GAGGGGAACCCTCTTTGGTTCCACCCTGAGCACCGAGCAGCCACTGCCCAGTACTTGT
CACCCCGGGCCAGGGATGCTGCTACTGGCTTCCTTCTCGATGGCAAGGTCTTGTCACT
GACAGATTTTCAGCAGACTCACACATCCTTGGGGCTCAGCCCCATGGGCCCACCTTTG
CCCTGGCCAGTGGGGAGTACTCCTGAAACCTCAGGTGGCCCTGACCTGAGTGACAGCC
TCTCCTCAGGGGGTGTTGTGACCCAGCCCCTGCTTCATAAGGTTAAGAGCCGAGTCCG
TGTGAGGCGGGCAAGCATCTCTGAACCCAGTGATACGGACCCGGAGCCCCGAACTCTG
AACCCCTCTCCGGCTGGTTGGTTCGTGCAGCAGCACCCGGAGCTGGAGCTCATGAGCA
GCTTCCGGGAACGGTTCGGCCGCAACTGGCTGCAGTACAGGAGTCACCTGGAGCCCTC
CGGAAACCCTCTGCCGGCCACCCCCACTACTTCTGCACCCAGTGCACCTCCAGCCAGC
TCCCAGGGCCCCGACACTGCACCCAGACCTTCACCCCCGCAGGAGGAAGCCAGAGGCC
CCCAGGAGTCACCACAGAAAATGTCAGAGGAGGTCAGGGCGGAGCCACAGGAGGAGGA
AGAGGAGAAGGAGGGGAAGGAGGAGAAGGAGGAGGGGGAGATGGTGGAACAGGGAGAA
GAGGAGGCAGGAGAGGAGGAAGAAGAGGAGCAGGACCAGAAGGAAGTGGAAGCGGAAC
TCTGTCGCCCCTTGTTGGTGTGTCCCCTGGAGGGGCCTGAGGGCGTACGGGGCAGGGA
ATGCTTTCTCAGGGTCACTTCTGCCCACCTGTTTGAGGTGGAACTCCAAGCAGCTCGC
ACCTTGGAGCGACTGGAGCTCCAGAGTCTGGAGGCAGCTGAGATAGAGCCGGAGGCCC
AGGCCCAGGGTCCCCCTCTTGCTGCGCAGGGCTCAGATCTGCTCCCTGGAGCCCCCAT
CCTCAGTCTGCGCTTCTCCTACATCTGCCCTGACCGGCAGTTGCGTCGCTATTTGGTG
CTGGAGCCTGATGCCCACGCAGCTGTCCAGGAGCTGCTTGCCGTGTTGACCCCAGTCA
CCAATGTGGCTCGGGAACAGCTTGGGGAGGCCAGGGACCTCCTGCTGGGTAGATTCCA
GTGTCTACGCTGTGGCCATGAGTTCAAGCCAGAGGAGCCCAGGATGGGATTAGACAGT
GAGGAAGGCTGGAGGCCTCTGTTCCAAAAGACAGAATCTCCTGCTGTGTGTCCTAACT
GTGGTAGTGACCACGTGGTTCTCCTCGCTGTGTCTCGGGGAACCCCCAACAGGGAGCG
GAAACAGGGAGAGCAGTCTCTGGCTCCTTCTCCGTCTGCCAGCCCTGTCTGCCACCCT
CCTGGCCATGGTGACCACCTTGACAGGGCCAAGAACAGCCCACCTCAGGCACCGAGCA
CCCGTGACCATGGTAGTTGGAGCCTCAGTCCCGCCCCTGAGCGCTGTGGCCTCCGCTC
TGTGGACCACCGACTCCGGCTCTTCCTGGATGTTGAGGTGTTCAGCGATGCCCAGGAG
GAGTTCCAGTGCTGCCTCAAGGTCCCAGTGGCATTGGCAGGCCACACTGGGGAGTTCA
TGTGCCTTGTGGTTGTGTCTGACCGCAGGCTGTACCTGTTGAAGGTGACTGGGGAGAT
GAGTGAGCCTCCAGCTAGCTGGCTGCAGCTGACCCTGGCTGTTCCCCTGCAGGATCTG
AGTGGCATAGAGCTGGGCCTGGCAGGCCAGAGCCTGCGGCTAGAGTGGGCAGCTGGGG
CGGGCCGCTGTGTGCTGCTGCCCCCGAGATGCCAGGCATTGCCGGGCCTTCCTAGAGGA
GCTCCTTGGTGTCTTGCAGTCTCTGCCCCCTGCCTGGAGGAACTGTGTCAGTGCCACA
GAGGAGGAGGTCACCCCCCAGCACCGGCTCTGGCCATTGCTGGAAAAAGACTCATCCT
TGGAGGCTCGCCAGTTCTTCTACCTTCGGGCGTTCCTGGTTGAAGGTGAAGCCTCTGT
GCAGCTGATGCTTCCCTCCACCTGCCTCGTATCCCTGTTGCTGACTCCGTCCACCCTG
TTCCTGTTAGATGAGGATGCTGCAGGGTCCCCGGCAGAGCCCTCTCCTCCAGCAGCAT
CTGGCGAAGCCTCTGAGAAGGTGCCTCCCTCGGGGCCGGGCCCTGCTGTGCGTGTCAG
GGAGCAGCAGCCACTCAGCAGCCTGAGCTCCGTGCTGCTCTACCGCTCAGCCCCTGAG
GACTTGCGGCTGCTCTTCTACGATGAGGTGTCCCGGCTGGAGAGCTTTTGGGCACTCC
GTGTGGTGTGTCAGGAGCAGCTGACAGCCCTGCTTGCCTGGATCCGGGAACCATGGGA
GGAGCTGTTTTCCATCGGACTCCGGACAGTGATCCAAGAGGCGCTGGCCCTTGACCGA
TGAGGGTCCCACGCTGACCTTGGCCCTGACCTCAGGAGCCACGCT</td></tr>
<tr><td></td><td>ORF Start: ATG at</td><td>ORF Stop: TGA at 3307</td></tr>
</table>

478

	13		
	SEQ ID NO: 354	1098 aa	MW at 121004.1kD
NOV127a, CG59971-01 Protein Sequence	MTTAQRDSLLWKLAGLLRESGDVVLSGCSTLSLLTPTLQQLNHVFELHLGPWGPGQTG FVALPSHPADSPVILQLQFLFDVLQKTLSLKLVHVAGPGPTGPIKIFPFKSLRHLELR GVPLHCLHGLRGIYSQLETLICSRSLQALEELLSACGGDFCSALPWLALLSANFSYNA LTALDSSLRLLSALRFLNLSHNQVQDCQGFLMDLCELHHLDISYNRLHLVPRMGPSGA ALGVLILRGNELRSLPGLEQLRNLRHLDLAYNLLEGHRELSPLWLLAELRKLYLEGNP LWFHPEHRAATAQYLSPRARDAATGFLLDGKVLSLTDFQQTHTSLGLSPMGPPLPWPV GSTPETSGGPDLSDSLSSGGVVTQPLLHKVKSRVRVRRASISEPSDTDPEPRTLNPSP AGWFVQQHPELELMSSFRERFGRNWLQYRSHLEPSGNPLPATPTTSAPSAPPASSQGP DTAPRPSPPQEEARGPQESPQKMSEEVRAEPQEEEEEKEGKEEKEEGEMVEQGEEEAG EEEEEEQDQKEVEAELCRPLLVCPLEGPEGVRGRECFLRVTSAHLFEVELQAARTLER LELQSLEAAEIEPEAQAQGPPLAAQGSDLLPGAPILSLRFSYICPDRQLRRYLVLEPD AHAAVQELLAVLTPVTNVAREQLGEARDLLLGRFQCLRCGHEFKPEEPRMGLDSEEGW RPLFQKTESPAVCPNCGSDHVVLLAVSRGTPNRERKQGEQSLAPSPSASPVCHPPGHG DHLDRAKNSPPQAPSTRDHGSWSLSPAPERCGLRSVDHRLRLFLDVEVFSDAQEEFQC CLKVPVALAGHTGEFMCLVVVSDRRLYLLKVTGEMSEPPASWLQLTLAVPLQDLSGIE LGLAGQSLRLEWAAGAGRCVLLPRDARHCRAFLEELLGVLQSLPPAWRNCVSATEEEV TPQHRLWPLLEKDSSLEARQFFYLRAFLVEGEASVQLMLPSTCLVSLLLTPSTLFLLD EDAAGSPAEPSPPAASGEASEKVPPSGPGPAVRVREQQPLSSLSSVLLYRSAPEDLRL LFYDEVSRLESFWALRVVCQEQLTALLAWIREPWEELFSIGLRTVIQEALALDR
	SEQ ID NO: 355	3348 bp	
NOV127b, CG59971-02 DNA Sequence	CGTCCCGTGGCC**ATG**ACGACCGCTCAGAGGGACTCCCTGTTGTGGAAGCTCGCGGGGT TGCTGCGGGAGTCCGGTGATGTGGTCCTGTCTGGCGTGTAGCACCCTGAGCCTGCTGAC TCCCACACTGCAACAGCTGAACCACGTATTTGAGCTGCACCTGGGGCCATGGGGCCCT GGCCAGACAGGCTTTGTGGCTCTGCCCTCCCATCCTGCCGACTCCCCTGTTATTCTTC AGCTTCAGTTTCTCTTCGATGTGCTGCAGAAAACACTTTCACTCAAGCTGGTCCATGT TGCTGGTCCTGGCCCCACAGGGCCCATCAAGATTTTCCCCTTCAAATCCCTTCGGCAC CTGGAGCTCCGAGGTGTTCCCCTCCACTGTCTGCATGGCCTCCGAGGCATCTACTCCC AGCTGGAGACCCTGATTTGCAGCAGGAGCCTCCAGGCATTAGAGGAGCTCCTCTCAGC CTGCGGCGGCGACTTCTGCTCTGCCCTCCCTTGGCTGGCTCTGCTTTCTGCCAACTTC AGCTACAATGCACTGACCGCCTTAGACAGCTCCCTGCGCCTCTTGTCAGCTCTGCGTT TCTTGAACCTAAGCCACAATCAAGTCCAGGACTGTCAGGGATTCCTGATGGATTTGTG TGAGCTCCACCATCTGGACATCTCCTATAATCGCCTGCATTTGGTGCCAAGAATGGGA CCCTCAGGGGCTGCTCTGGGGGTCCTGATACTGCGAGGCAATGAGCTTCGGAGCCTGC CAGGCCTAGAGCAGCTGAGGAATCTGCGGCACCTGGATTTGGCATACAACCTGCTGGA AGGACACCGGGAGCTGTCACCACTGTGGCTGCTGGCTGAGCTCCGCAAGCTCTACCTG GAGGGGAACCCTCTTTGGTTCCACCCTGAGCACCGAGCAGCCACTGCCCAGTACTTGT CACCCCGGGCCAGGGATGCTGCTACTGGCTTCCTTCTCGATGGCAAGGTCTTGTCACT GACAGATTTTCAGCAGACTCACACATCCTTGGGGCTCAGCCCCATGGGCCCCACCTTTG CCCTGGCCAGTGGGGAGTACTCCTGAAACCTCAGGTGGCCCTGACCTGAGTGACAGCC TCTCCTCAGGGGGTGTTGTGACCCAGCCCCTGCTTCATAAGGTTAAGAGCCGAGTCCG TGTGAGGCGGGCAAGCATCTCTGAACCCAGTGATACGGACCCGGAGCCCCGAACTCTG AACCCCTCTCCGGCTGGTTGGTTCGTGCAGCAGCACCCGGAGCTGGAGCTCATGAGCA GCTTCCGGGAACGGTTCGGCCGCAACTGGCTGCAGTACAGGAGTCACCTGGAGCCCTC CGGAAACCCTCTGCCGGCCACCCCCACTACTTCTGCACCCAGTGCACCTCCAGCCAGC TCCCAGGGCCCCGACACTGCACCCAGACCTTCACCCCCGCAGGAGGAAGCCAGAGGCC CCCAGGAGTCACCACAGAAAATGTCAGAGGAGGTCAGGGCGGAGCCACAGGAGGAGGA AGAGGAGAAGGAGGGGAAGGAGGAGAAGGAGGAGGGGGAGATGGTGGAACAGGGAGAA GAGGAGGCAGGAGAGGAGGAAGAAGAGGAGCAGGACCAGAAGGAAGTGGAAGCGGAAC TCTGTCGCCCCTTGTTGGTGTGTCCCCTGGAGGGGCCTGAGGGCGTACGGGGCAGGGA ATGCTTTCTCAGGGTCACTTCTGCCCACCTGTTTGAGGTGGAACTCCAAGCAGCTCGC ACCTTGGAGCGACTGGAGCTCCAGAGTCTGGAGGCAGCTGAGATAGAGCCGGAGGCCC AGGCCCAGAGGTCGCCCAGGCCCACGGGCTCAGATCTGCTCCCTGGAGCCCCCATCCT CAGTCTGCGCTTCTCCTACATCTGCCCTGACCGGCAGTTGCGTCGCTATTTGGTGCTG GAGCCTGATGCCCACGCAGCTGTCCAGGAGCTGCTTGCCGTGTTGACCCCAGTCACCA ATGTGGCTCGGGAACAGCTTGGGGAGGCCAGGGACCTCCTGCTGGGTAGATTCCAGTG TCTACGCTGTGGCCATGAGTTCAAGCCAGAGGAGCCCAGGATGGGATTAGACAGTGAG GAAGGCTGGAGGCCTCTGTTCCAAAAGACAGAATCTCCTGCTGTGTGTCCTAACTGTG GTAGTGACCACGTGGTTCTCCTCGCTGTGTCTCGGGGAACCCCCAACAGGGAGCGGAA ACAGGGAGAGCAGTCTCTGGCTCCTTCTCCGTCTGCCAGCCCTGTCTGCCACCCTCCT GGCCATGGTGACCACCTTGACAGGGCCAAGAACAGCCCACCTCAGGCACCGAGCACCC GTGACCATGGTAGTTGGAGCCTCAGTCCCGCCCCTGAGCGCTGTGGCCTCCGCTCTGT GGACCACCGACTCCGGCTCTTCCTGGATGTTGAGGTGTTCAGCGATGCCCAGGAGGAG TTCCAGTGCTGCCTCAAGGTCCCAGTGGCATTGGCAGGCCACACTGGGGAGTTCATGT GCCTTGTGGTTGTGTCTGACCGCAGGCTGTACCTGTTGAAGGTGACTGGGGAGATGAG TGAGCCTCCAGCTAGCTGGCTGCAGCTGACCCTGGCTGTTCCCCTGCAGGATCTGAGT GGCATAGAGCTGGGCCTGGCAGGCCAGAGCCTGCGGCTAGAGTGGGCAGCTGGGGCGG GCCGCTGTGTGCTGCTGCCCCGAGATGCCAGGCATTGCCGGGCCTTCCTAGAGGAGCT CCTTGGTGTCTTGCAGTCTCTGCCCCCTGCCTGGAGGAACTGTGTCAGTGCCACAGAG GAGGAGGTCACCCCCCAGCACCGGCTCTGGCCATTGCTGGAAAAAGACTCATCCTTGG AGGCTCGCCAGTTCTTCTACCTTGGGCGTTCCTGGTTGAAGGTGAAGCCTCTGTGCA GCTGATGCTTCCCTCCACCTGCCTCGTATCCCTGTTGCTGACTCCGTCCACCCTGTTC

479

	CTGTTAGATGAGGATGCTGCAGGGTCCCCGGCAGAGCCCTCTCCTCCAGCAGCATCTG GCGAAGCCTCTGAGAAGGTGCCTCCCTCGGGGCCGGGCCCTGCTGTGCGTGTCAGGGA GCAGCAGCCACTCAGCAGCCTGAGCTCCGTGCTGCTCTACCGCTCAGCCCCTGAGGAC TTGCGGCTGCTCTTCTACGATGAGGTGTCCCGGCTGGAGAGCTTTTGGGCACTCCGTG TGGTGTGTCAGGAGCAGCTGACAGCCCTGCTTGCCTGGATCCGGGAACCATGGGAGGA GCTGTTTTCCATCGGACTCCGGACAGTGATCCAAGAGGCGCTGGCCCTTGACCGA**TGA** GGGTCCCACGCTGACCTTGGCCCTGACCTCAGGAGCCACGCT
	ORF Start: ATG at 13	ORF Stop: TGA at 3304	
	SEQ ID NO: 356	1097 aa	MW at 121064.1kD
NOV127b, CG59971-02 Protein Sequence	MTTAQRDSLLWKLAGLLRESGDVVLSGCSTLSLLTPTLQQLNHVFELHLGPWGPGQTG FVALPSHPADSPVILQLQFLFDVLQKTLSLKLVHVAGPGPTGPIKIFPFKSLRHLELR GVPLHCLHGLRGIYSQLETLICSRSLQALEELLSACGGDFCSALPWLALLSANFSYNA LTALDSSLRLLSALRFLNLSHNQVQDCQGFLMDLCELHHLDISYNRLHLVPRMGPSGA ALGVLILRGNELRSLPGLEQLRNLRHLDLAYNLLEGHRELSPLWLLAELRKLYLEGNP LWFHPEHRAATAQYLSPRARDAATGFLLDGKVLSLTDFQQTHTSLGLSPMGPPLPWPV GSTPETSGGPDLSDSLSSGGVVTQPLLHKVKSRVRVRRASISEPSDTDPEPRTLNPSP AGWFVQQHPELELMSSFRERFGRNWLQYRSHLEPSGNPLPATPTTSAPSAPPASSQGP DTAPRPSPPQEEARGPQESPQKMSEEVRAEPQEEEEEKEGKEEKEEGEMVEQGEEEAG EEEEEEQDQKEVEAELCRPLLVCPLEGPEGVRGRECFLRVTSAHLFEVELQAARTLER LELQSLEAAEIEPEAQAQRSPRPTGSDLLPGAPILSLRFSYICPDRQLRRYLVLEPDA HAAVQELLAVLTPVTNVAREQLGEARDLLLGRFQCLRCGHEFKPEEPRMGLDSEEGWR PLFQKTESPAVCPNCGSDHVVLLAVSRGTPNRERKQGEQSLAPSPSASPVCHPPGHGD HLDRAKNSPPQAPSTRDHGSWSLSPAPERCGLRSVDHRLRLFLDVEVFSDAQEEFQCC LKVPVALAGHTGEFMCLVVVSDRRLYLLKVTGEMSEPPASWLQLTLAVPLQDLSGIEL GLAGQSLRLEWAAGAGRCVLLPRDARHCRAFLEELLGVLQSLPPAWRNCVSATEEEVT PQHRLWPLLEKDSSLEARQFFYLRAFLVEGEASVQLMLPSTCLVSLLLTPSTLFLLDE DAAGSPAEPSPPAASGEASEKVPPSGPGPAVRVREQQPLSSLSSVLLYRSAPEDLRLL FYDEVSRLESFWALRVVCQEQLTALLAWIREPWEELFSIGLRTVIQEALALDR

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 127B.

Table 127B. Comparison of NOV127a against NOV127b.		
Protein Sequence	NOV127a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV127b	1..1098 1..1097	891/1098 (81%) 891/1098 (81%)

Further analysis of the NOV127a protein yielded the following properties shown in Table 127C.

Table 127C. Protein Sequence Properties NOV127a	
PSort analysis:	0.5163 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.2442 probability located in mitochondrial inner membrane; 0.2442 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

480

A search of the NOV127a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 127D.

Table 127D. Geneseq Results for NOV127a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV127a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM39827	Human polypeptide SEQ ID NO 2972 - Homo sapiens, 169 aa. [WO200153312-A1, 26-JUL-2001]	375..528 14..167	140/154 (90%) 145/154 (93%)	3e-78
AAM41613	Human polypeptide SEQ ID NO 6544 - Homo sapiens, 184 aa. [WO200153312-A1, 26-JUL-2001]	375..528 29..182	140/154 (90%) 145/154 (93%)	4e-78
AAU19764	Human novel extracellular matrix protein, Seq ID No 414 - Homo sapiens, 211 aa. [WO200155368-A1, 02-AUG-2001]	444..647 13..209	157/207 (75%) 160/207 (76%)	2e-75
ABB19833	Protein #1832 encoded by probe for measuring heart cell gene expression - Homo sapiens, 127 aa. [WO200157274-A2, 09-AUG-2001]	409..535 1..127	127/127 (100%) 127/127 (100%)	2e-70
AAM67606	Human bone marrow expressed probe encoded protein SEQ ID NO: 27912 - Homo sapiens, 127 aa. [WO200157276-A2, 09-AUG-2001]	409..535 1..127	127/127 (100%) 127/127 (100%)	2e-70

In a BLAST search of public sequence databases, the NOV127a protein was found to have homology to the proteins shown in the BLASTP data in Table 127E.

Table 127E. Public BLASTP Results for NOV127a				
Protein Accession Number	Protein/Organism/Length	NOV127a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAL49726	LKB1-INTERACTING PROTEIN 1 - Homo sapiens (Human), 1099 aa.	1..1098 12..1099	1077/1098 (98%) 1078/1098 (98%)	0.0

481

Q96PY9	KIAA1898 PROTEIN - Homo sapiens (Human), 1013 aa (fragment).	76..1098 1..1013	1003/1023 (98%) 1003/1023 (98%)	0.0
Q96CN3	SIMILAR TO RIKEN CDNA 1200014D22 GENE - Homo sapiens (Human), 804 aa (fragment).	288..1098 4..804	793/811 (97%) 793/811 (97%)	0.0
Q9DBT7	1200014D22RIK PROTEIN - Mus musculus (Mouse), 1072 aa.	1..1098 1..1072	816/1098 (74%) 895/1098 (81%)	0.0
Q9VMK9	CG9044 PROTEIN - Drosophila melanogaster (Fruit fly), 1289 aa.	12..433 8..463	139/459 (30%) 220/459 (47%)	6e-38

PFam analysis predicts that the NOV127a protein contains the domains shown in the Table 127F.

Table 127F. Domain Analysis of NOV127a			
Pfam Domain	NOV127a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRR: domain 1 of 5	164..186	7/25 (28%) 15/25 (60%)	2.5e+02
LRR: domain 2 of 5	187..209	6/25 (24%) 16/25 (64%)	2.5e+02
LRR: domain 3 of 5	210..231	8/25 (32%) 13/25 (52%)	83
LRR: domain 4 of 5	233..254	9/25 (36%) 17/25 (68%)	16
LRR: domain 5 of 5	255..279	10/27 (37%) 19/27 (70%)	22
Pkinase_C: domain 1 of 1	620..629	5/11 (45%) 9/11 (82%)	8.9
rubredoxin: domain 1 of 2	669..686	5/18 (28%) 13/18 (72%)	4.6
rubredoxin: domain 2 of 2	708..713	5/6 (83%) 6/6 (100%)	1.2e+03

5

Example B: Sequencing Methodology and Identofication of NOVX Clones

1.   **GeneCalling™ Technology:** This is a proprietary method of performing differential gene expression profiling between two or more samples developed at CuraGen and described by Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999). cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

2.   **SeqCalling™ Technology:** cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included

483

in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other

5   sequence variations.

3.   PathCalling™ Technology:

The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the

10   DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

15   cDNA libraries were derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors,

20   chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in

25   their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corportion proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain

30   reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for

corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the

5 extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering

10 the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

15

4. **RACE:** Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue

20 samples from different donors were used for the RACE reaction. The sequences derived from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5. **Exon Linking:** The NOVX target sequences identified in the present invention

25 were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. **Table B1** shows the sequences of the PCR primers used for obtaining different clones. In each case, the sequence was examined, walking inward from the respective termini toward the coding

30 sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated

homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal

5   kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial

10   clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were

15   evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

6.   Physical Clone:

Exons were predicted by homology and the intron/exon boundaries were

20   determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent

25   inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

30

Example C: Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines

35   and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an

Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on metabolic diseases), AI_comprehensive_panel (containing normal tissue and samples from autoimmune diseases), Panel CNSD.01 (containing central nervous system samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β-actin and GAPDH). Normalized RNA (5 ul) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10 μg of total RNA were performed in a volume of 20 μl and incubated for 60 minutes at 42°C. This reaction can be scaled up to 50 μg of total RNA in a final volume of 100 μl. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC)

or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (Tm) range = 58°-60°C, primer optimal Tm = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe Tm

5   must be 10°C greater than primer Tm, amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM

10   each, and probe, 200nM.

    PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone

15   and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were recorded as CT values

20   (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

25   When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15

30   seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

Panels 1, 1.1, 1.2, and 1.3D

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,
* = established from metastasis,
met = metastasis,
s cell var = small cell variant,
non-s = non-sm = non-small,
squam = squamous,
pl. eff = pl effusion = pleural effusion,
glio = glioma,
astro = astrocytoma, and
neuro = neuroblastoma.

General_screening_panel_v1.4

The plates for Panel 1.4 include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panel 1.4 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS

489

cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panel 1.4 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on

5   Panel 1.4 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland,

10   spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

Panels 2D and 2.2

         The plates for Panels 2D and 2.2 generally include 2 control wells and 94 test

15   samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor.

20   These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding

25   the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and

30   were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

490

Panel 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines

5   are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there

10   are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

Panels 4D, 4R, and 4.1D

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples)

15   composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal

20   tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial

25   cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-

30   5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately

5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2μg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately 2×10^6cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol (5.5×10^{-5}M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10μg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56,

5   CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}M$ (Gibco), and

10   10mM Hepes (Gibco) and plated at 10^6cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5μg/ml anti-CD28 (Pharmingen) and 3ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested

15   the cells and expanded them in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}M$ (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture.

20   The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}M$ (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun

25   down and resupended at 10^6cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}M$ (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5μg/ml or anti-CD40 (Pharmingen) at approximately 10μg/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

30   To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10μg/ml anti-CD28 (Pharmingen) and 2μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^5-10^6cells/ml in DMEM 5% FCS

493

(Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1μg/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1μg/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to

5   Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as

10   described above, but with the addition of anti-CD95L (1μg/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-

15   CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at 5×10^5cells/ml for 8 days, changing the media every 3 days and adjusting the cell

20   concentration to 5×10^5cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1μg/ml for 6 and 14 hours. Keratinocyte

25   line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were

30   activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane

(Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20°C overnight. The precipitated RNA was spun down

5   at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300μl of RNAse-free water and 35μl buffer (Promega) 5μl DTT, 7μl RNAsin and 8μl DNAse were added. The tube was incubated at 37°C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The

10   RNA was spun down and placed in RNAse free water. RNA was stored at -80°C.

AI_comprehensive panel_v1.0

The plates for AI_comprehensive panel_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was

15   extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated

20   RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the

25   ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used.

30   Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and

four female patients. Four of the patients were taking lebvid and two were on phenobarbital.

Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients

5   ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids,

10   and bronchodilators.

In the labels employed to identify tissues in the AI_comprehensive panel_v1.0 panel, the following abbreviations are used:

AI = Autoimmunity
Syn = Synovial
15   Normal = No apparent disease
Rep22 /Rep20 = individual patients
RA = Rheumatoid arthritis
Backus = From Backus Hospital
OA = Osteoarthritis
20   (SS) (BA) (MF) = Individual patients
Adj = Adjacent tissue
Match control = adjacent tissues
-M = Male
-F = Female
25   COPD = Chronic obstructive pulmonary disease

Panels 5D and 5I

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study.

30   Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being

496

repaired/closed, the obstetrician removed a small sample (<1 cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

Patient 2	Diabetic Hispanic, overweight, not on insulin
Patient 7-9	Nondiabetic Caucasian and obese (BMI>30)
Patient 10	Diabetic Hispanic, overweight, on insulin
Patient 11	Nondiabetic African American and overweight
Patient 12	Diabetic Hispanic on insulin

Adipocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose
Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated
Donor 2 and 3 AD: Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the

497

University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

5   GO Adipose = Greater Omentum Adipose
           SK = Skeletal Muscle
           UT = Uterus
           PL = Placenta
           AD = Adipose Differentiated
10   AM = Adipose Midway Differentiated
           U = Undifferentiated Stem Cells

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the
15   Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains
20   from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area
25   17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to
30   be free of any pathology consistent with neurodegeneration.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

498

PSP = Progressive supranuclear palsy
Sub Nigra = Substantia nigra
Glob Palladus= Globus palladus
Temp Pole = Temporal pole
5   Cing Gyr = Cingulate gyrus
BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

The plates for Panel CNS_Neurodegeneration_V1.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained

10   from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

15   Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like

20   pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a

25   region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

30   In the labels employed to identify tissues in the CNS_Neurodegeneration_V1.0 panel, the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy

499

Control = Control brains; patient not demented, showing no neuropathology
Control (Path) = Control brains; pateint not demented but showing sever AD-like pathology
SupTemporal Ctx = Superior Temporal Cortex
Inf Temporal Ctx = Inferior Temporal Cortex

A. CG58522-01: HUMAN PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE IB BETA

Expression of gene CG58522-01 was assessed using the primer-probe set Ag3365, described in Table AA. Results of the RTQ-PCR runs are shown in Table AB.

Table AA. Probe Name Ag3365

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cagaatgaaccaaggagactca-3'	22	3	357
Probe	TET-5'-ctactccgcatgcggcagaagacatt-3'-TAMRA	26	35	358
Reverse	5'-cacatccatctgtcatctcctt-3'	22	62	359

Table AB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3365, Run 216709759	Tissue Name	Rel. Exp.(%) Ag3365, Run 216709759
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	10.7	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca.	0.0	Colon ca. Colo-205	0.0

OVCAR-3			
Ovarian ca. SK-OV-3	4.9	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	7.9	Stomach Pool	0.0
Ovarian ca. OVCAR-8	26.8	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	1.7	Lymph Node Pool	16.5
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	3.3	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	4.5	CNS cancer (glio) SNB-19	6.2
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	25.7
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	**100.0**	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	4.8
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	1.8
Liver	0.0	Brain (Thalamus) Pool	3.6
Fetal Liver	0.0	Brain (whole)	6.9
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0

Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

CNS_neurodegeneration_v1.0 Summary: Ag3365 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3365 - Significant expression of this gene is seen only in the lung cancer cell line NCI-H23 (CT=33.1). Therefore, expression of this gene may be used to distinguish this sample from the other samples on this panel.

Panel 4D Summary: Ag3365 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

B. CG58520-01: GAMMA-AMINOBUTYRIC-ACID RECEPTOR GAMMA-1

Expression of gene CG58520-01 was assessed using the primer-probe set Ag3364, described in Table BA.

Table BA. Probe Name Ag3364

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ttcttctgcggagtcaaagtag-3'	22	43	360
Probe	TET-5'-ttggtcttcttgttactgaccctgca-3'-TAMRA	26	75	361
Reverse	5'-tcatctgccttatcaacgtttc-3'	22	106	362

CNS_neurodegeneration_v1.0 Summary: Ag3364 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3364 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag3364 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel CNS_1 Summary: Ag3364 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

C. CG58520-03: GAMMA-AMINOBUTYRIC-ACID RECEPTOR GAMMA-1 SUBUNIT PRECURSOR (GABA(A) RECEPTOR)

Expression of gene CG58520-03 was assessed using the primer-probe set Ag5092, described in Table CA.

Table CA. Probe Name Ag5092

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaacattcctgtccactgga-3'	20	625	363
Probe	TET-5'-attttcaagcgatggataccctaaaa-3'-TAMRA	26	645	364
Reverse	5'-cacttctacggagggctttt-3'	20	692	365

CNS_neurodegeneration_v1.0 Summary: Ag5092 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.5 Summary: Ag5092 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag5092 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

D. CG58518-01: GAMMA-AMINOBUTYRIC-ACID RECEPTOR RHO-3 -

Expression of gene CG58518-01 was assessed using the primer-probe sets Ag3363, Ag1130, Ag1198, Ag1253 and Ag1603, described in Tables DA, DB, DC, DD and DE. Results of the RTQ-PCR runs are shown in Tables DF, DG and DH.

Table DA. Probe Name Ag3363

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tggctttccagttagtctcctt-3'	22	14	366
Probe	TET-5'-cacctacatctggatcatattgaaacca-3'-TAMRA	28	36	367
Reverse	5'-ttgatgttagaagcagcacaaa-3'	22	68	368

503

Table DB. Probe Name Ag1130

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtcctggctttccagttagtct-3'	22	10	369
Probe	TET-5'-tcacctacatctggatcatattgaaacca-3'-TAMRA	29	35	370
Reverse	5'-ttgatgttagaagcagcacaaa-3'	22	68	371

Table DC. Probe Name Ag1198

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtcctggctttccagttagtct-3'	22	10	372
Probe	TET-5'-tcacctacatctggatcatattgaaacca-3'-TAMRA	29	35	373
Reverse	5'-ttgatgttagaagcagcacaaa-3'	22	68	374

Table DD. Probe Name Ag1253

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atctgggtgcctgatatctttt-3'	22	466	375
Probe	TET-5'-tgtccactctaaaagatccttcatccatga-3'-TAMRA	30	489	376
Reverse	5'-cgcagcatgatattctccatag-3'	22	524	377

Table DE. Probe Name Ag1603

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtcctggctttccagttagtct-3'	22	10	378
Probe	TET-5'-tcacctacatctggatcatattgaaacca-3'-TAMRA	29	35	379
Reverse	5'-ttgatgttagaagcagcacaaa-3'	22	68	380

Table DF. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3363, Run 216709559	Tissue Name	Rel. Exp.(%) Ag3363, Run 216709559
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	6.6
Melanoma*	0.0	Gastric ca. (liver met.)	0.0

504

Hs688(B).T		NCI-N87	
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	16.7	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	6.8
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	6.4	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	8.5
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	10.9
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0

505

Lung ca. NCI-H146	77.9	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	**100.0**	CNS cancer (glio) SF-295	11.4
Lung ca. A549	10.1	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	34.4	Brain (fetal)	0.0
Lung ca. NCI-H460	30.6	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	5.1
Fetal Liver	0.0	Brain (whole)	50.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	3.0	Adrenal Gland	0.0
Fetal Kidney	8.4	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table DG. Panel 1.2

Tissue Name	Rel. Exp.(%) Ag1130, Run 125117140	Rel. Exp.(%) Ag1130, Run 126566764	Rel. Exp.(%) Ag1198, Run 129140506	Tissue Name	Rel. Exp.(%) Ag1130, Run 125117140	Rel. Exp.(%) Ag1130, Run 126566764	Rel. Exp.(%) Ag1198, Run 129140506
Endothelial cells	0.0	0.0	0.0	Renal ca. 786-0	0.0	0.0	0.0
Heart (Fetal)	0.0	0.0	0.0	Renal ca. A498	7.3	4.7	0.0
Pancreas	0.0	0.0	0.0	Renal ca. RXF 393	0.0	0.0	0.0
Pancreatic ca. CAPAN 2	9.0	0.0	0.0	Renal ca. ACHN	0.0	0.0	0.0
Adrenal Gland	0.0	2.6	0.0	Renal ca. UO-31	3.9	0.0	0.0
Thyroid	0.0	0.0	0.0	Renal ca. TK-10	0.0	0.0	0.0
Salivary gland	0.0	0.0	0.0	Liver	26.6	0.0	0.0

506

Pituitary gland	0.0	0.0	0.0	Liver (fetal)	25.3	0.0	0.0
Brain (fetal)	0.0	0.0	0.0	Liver ca. (hepatobla st) HepG2	0.0	0.0	0.0
Brain (whole)	2.6	20.0	0.0	Lung	0.0	0.0	0.0
Brain (amygdala)	1.3	32.1	0.0	Lung (fetal)	0.0	0.0	0.0
Brain (cerebellum)	1.5	3.8	0.0	Lung ca. (small cell) LX-1	3.4	0.0	0.0
Brain (hippocamp us)	0.0	27.0	0.0	Lung ca. (small cell) NCI-H69	28.5	74.2	0.0
Brain (thalamus)	9.9	22.5	9.8	Lung ca. (s.cell var.) SHP-77	3.8	9.7	0.0
Cerebral Cortex	0.0	0.0	0.0	Lung ca. (large cell)NCI-H460	8.8	4.1	5.3
Spinal cord	4.4	0.0	0.0	Lung ca. (non-sm. cell) A549	51.4	9.5	7.2
glio/astro U87-MG	0.0	0.0	0.0	Lung ca. (non-s.cell) NCI-H23	0.0	0.0	0.0
glio/astro U-118-MG	0.0	0.0	0.0	Lung ca. (non-s.cell) HOP-62	8.4	2.7	9.6
astrocytom a SW1783	2.9	0.0	0.0	Lung ca. (non-s.cl) NCI-H522	0.0	0.0	0.0
neuro*; met SK-N-AS	0.0	0.0	0.0	Lung ca. (squam.) SW 900	3.2	8.7	0.0
astrocytom a SF-539	5.1	0.0	0.0	Lung ca. (squam.) NCI-H596	2.3	15.9	0.0
astrocytom	2.3	0.0	0.0	Mammary	0.0	0.0	0.0

507

a SNB-75				gland			
glioma SNB-19	6.3	20.7	9.0	Breast ca.* (pl.ef) MCF-7	0.0	0.0	0.0
glioma U251	1.4	0.0	1.8	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0	0.0
glioma SF-295	0.0	0.0	0.0	Breast ca.* (pl. ef) T47D	14.1	37.4	0.0
Heart	0.0	0.0	0.0	Breast ca. BT-549	12.5	21.0	12.3
Skeletal Muscle	2.3	0.0	0.0	Breast ca. MDA-N	0.0	0.0	0.0
Bone marrow	0.0	0.0	0.0	Ovary	0.0	0.0	0.0
Thymus	0.0	0.0	0.0	Ovarian ca. OVCAR-3	0.0	0.0	0.0
Spleen	2.2	0.0	0.0	Ovarian ca. OVCAR-4	0.0	0.0	0.0
Lymph node	0.0	0.0	0.0	Ovarian ca. OVCAR-5	66.9	35.4	4.4
Colorectal Tissue	11.3	27.7	21.8	Ovarian ca. OVCAR-8	2.7	0.0	0.0
Stomach	0.0	0.0	0.0	Ovarian ca. IGROV-1	6.0	0.0	0.0
Small intestine	5.4	0.0	0.0	Ovarian ca. (ascites) SK-OV-3	30.8	0.0	0.0
Colon ca. SW480	3.2	0.0	0.0	Uterus	0.0	0.0	0.0
Colon ca.* SW620 (SW480	0.0	0.0	0.0	Placenta	0.0	0.0	0.0

met)							
Colon ca. HT29	1.9	14.4	0.0	Prostate	6.9	0.0	0.0
Colon ca. HCT-116	0.0	0.0	0.0	Prostate ca.* (bone met) PC-3	**100.0**	0.0	0.0
Colon ca. CaCo-2	0.0	0.0	0.0	Testis	54.7	**100.0**	36.9
Colon ca. Tissue (ODO3866)	72.2	75.8	**100.0**	Melanoma Hs688(A).T	4.2	0.0	0.0
Colon ca. HCC-2998	5.3	4.8	0.0	Melanoma* (met) Hs688(B).T	2.7	34.2	13.3
Gastric ca.* (liver met) NCI-N87	50.3	0.0	0.0	Melanoma UACC-62	0.0	0.0	0.0
Bladder	6.0	22.1	0.0	Melanoma M14	31.4	36.3	20.2
Trachea	0.0	0.0	0.0	Melanoma LOX IMVI	0.0	0.0	0.0
Kidney	2.0	0.0	0.0	Melanoma* (met) SK-MEL-5	2.4	0.0	0.0
Kidney (fetal)	1.1	2.5	0.0				

Table DH. Panel 4R

Tissue Name	Rel. Exp.(%) Ag1198, Run 142014937	Tissue Name	Rel. Exp.(%) Ag1198, Run 142014937
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	2.5	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC	0.0

509

		TNFalpha + IL-1beta	
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	16.4
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell)	0.0	Lung fibroblast IL-13	0.0

510

ionomycin			
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 1	**100.0**
Monocytes rest	0.0	IBD Colitis 2	0.0
Monocytes LPS	0.0	IBD Crohn's	0.0
Macrophages rest	0.0	Colon	0.0
Macrophages LPS	0.0	Lung	0.0
HUVEC none	0.0	Thymus	0.0
HUVEC starved	0.0	Kidney	0.0

CNS_neurodegeneration_v1.0 Summary: Ag3363 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3363 - Significant expression is seen in lung cancer cell line NCI-H146 (CT=34.5) and lung cancer cell line SHP-77 (CT=34.2). Therefore, expression of this can be used to distinguish these samples from the rest of the samples on this panel.

Panel 1.2 Summary: Ag1130/Ag1198 - Three different runs using the same primer sequences yield similar results. Significant expression of this gene is seen in testis and a colon cancer sample. Therefore, expression of this gene can be used to differentiate these samples from other samples on these panels. Results from a third experiment using the probe and primer set Ag1253 show low/undetectable levels of expression in all the samples on this panel.

Panel 1.3D Summary: Ag1253 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 2D Summary: Ag1603 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag1130/Ag1198/Ag1253/Ag3363 - Two experiments showed possible experimental difficulties, while the other three runs showed expression of this gene as low/undetectable (CTs > 35) across all of the samples on the panel.

Panel 4R Summary: Ag1198 - Significant expression of this gene is seen only in the IBD colitis 1 sample (CT=34.2). Therefore, expression of this gene can be used to differentiate this sample from others on the panel.

Panel CNS_1 Summary: Ag1253/Ag1603 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

E. CG58516-01: G-protein beta WD-40 repeats

Expression of gene CG58516-01 was assessed using the primer-probe set Ag3362, described in Table EA. Results of the RTQ-PCR runs are shown in Tables EB and EC.

Table EA. Probe Name Ag3362

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtcgggcaggacctttact-3'	19	1474	381
Probe	TET-5'-tcctacagctaattctgcagggcaca-3'-TAMRA	26	1498	382
Reverse	5'-tacgctttactcccgtaagtca-3'	22	1543	383

Table EB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3362, Run 210153738	Tissue Name	Rel. Exp.(%) Ag3362, Run 210153738
AD 1 Hippo	9.9	Control (Path) 3 Temporal Ctx	0.0
AD 2 Hippo	33.2	Control (Path) 4 Temporal Ctx	24.3
AD 3 Hippo	4.3	AD 1 Occipital Ctx	2.0
AD 4 Hippo	16.5	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	96.6	AD 3 Occipital Ctx	5.4

512

AD 6 Hippo	43.2	AD 4 Occipital Ctx	24.7
Control 2 Hippo	29.1	AD 5 Occipital Ctx	24.5
Control 4 Hippo	16.6	AD 6 Occipital Ctx	31.9
Control (Path) 3 Hippo	3.8	Control 1 Occipital Ctx	0.9
AD 1 Temporal Ctx	7.1	Control 2 Occipital Ctx	89.5
AD 2 Temporal Ctx	23.2	Control 3 Occipital Ctx	12.6
AD 3 Temporal Ctx	5.6	Control 4 Occipital Ctx	6.3
AD 4 Temporal Ctx	20.0	Control (Path) 1 Occipital Ctx	65.1
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	15.8
AD 5 SupTemporal Ctx	43.8	Control (Path) 3 Occipital Ctx	2.0
AD 6 Inf Temporal Ctx	30.8	Control (Path) 4 Occipital Ctx	11.6
AD 6 Sup Temporal Ctx	69.7	Control 1 Parietal Ctx	2.8
Control 1 Temporal Ctx	9.0	Control 2 Parietal Ctx	39.2
Control 2 Temporal Ctx	59.0	Control 3 Parietal Ctx	23.5
Control 3 Temporal Ctx	11.7	Control (Path) 1 Parietal Ctx	69.7
Control 4 Temporal Ctx	8.2	Control (Path) 2 Parietal Ctx	14.9
Control (Path) 1 Temporal Ctx	56.3	Control (Path) 3 Parietal Ctx	0.9
Control (Path) 2 Temporal Ctx	34.2	Control (Path) 4 Parietal Ctx	38.7

Table EC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3362, Run 216523482	Tissue Name	Rel. Exp.(%) Ag3362, Run 216523482
Adipose	6.3	Renal ca. TK-10	44.1
Melanoma* Hs688(A).T	17.6	Bladder	9.4

513

Melanoma* Hs688(B).T	18.3	Gastric ca. (liver met.) NCI-N87	21.6
Melanoma* M14	17.1	Gastric ca. KATO III	17.6
Melanoma* LOXIMVI	13.6	Colon ca. SW-948	5.8
Melanoma* SK-MEL-5	19.6	Colon ca. SW480	34.6
Squamous cell carcinoma SCC-4	14.6	Colon ca.* (SW480 met) SW620	14.2
Testis Pool	4.0	Colon ca. HT29	7.2
Prostate ca.* (bone met) PC-3	90.8	Colon ca. HCT-116	14.3
Prostate Pool	4.0	Colon ca. CaCo-2	19.8
Placenta	11.4	Colon cancer tissue	3.6
Uterus Pool	2.1	Colon ca. SW1116	9.4
Ovarian ca. OVCAR-3	17.4	Colon ca. Colo-205	8.8
Ovarian ca. SK-OV-3	47.0	Colon ca. SW-48	13.2
Ovarian ca. OVCAR-4	14.7	Colon Pool	5.7
Ovarian ca. OVCAR-5	31.6	Small Intestine Pool	10.2
Ovarian ca. IGROV-1	12.9	Stomach Pool	6.2
Ovarian ca. OVCAR-8	6.7	Bone Marrow Pool	1.3
Ovary	12.5	Fetal Heart	1.1
Breast ca. MCF-7	75.8	Heart Pool	3.4
Breast ca. MDA-MB-231	30.4	Lymph Node Pool	8.7
Breast ca. BT 549	65.5	Fetal Skeletal Muscle	2.3
Breast ca. T47D	**100.0**	Skeletal Muscle Pool	9.4
Breast ca. MDA-N	33.4	Spleen Pool	4.6
Breast Pool	4.6	Thymus Pool	7.3
Trachea	7.7	CNS cancer (glio/astro) U87-MG	33.9
Lung	4.9	CNS cancer (glio/astro) U-118-MG	27.2
Fetal Lung	7.1	CNS cancer (neuro;met) SK-N-AS	16.0
Lung ca. NCI-N417	9.3	CNS cancer (astro) SF-539	14.3
Lung ca. LX-1	15.8	CNS cancer (astro) SNB-75	60.7

Lung ca. NCI-H146	4.9	CNS cancer (glio) SNB-19	13.8
Lung ca. SHP-77	16.5	CNS cancer (glio) SF-295	28.5
Lung ca. A549	27.2	Brain (Amygdala) Pool	5.3
Lung ca. NCI-H526	4.1	Brain (cerebellum)	5.0
Lung ca. NCI-H23	15.0	Brain (fetal)	16.4
Lung ca. NCI-H460	9.5	Brain (Hippocampus) Pool	5.5
Lung ca. HOP-62	7.6	Cerebral Cortex Pool	8.7
Lung ca. NCI-H522	18.2	Brain (Substantia nigra) Pool	8.3
Liver	0.0	Brain (Thalamus) Pool	6.3
Fetal Liver	7.3	Brain (whole)	7.0
Liver ca. HepG2	29.5	Spinal Cord Pool	5.6
Kidney Pool	17.7	Adrenal Gland	6.3
Fetal Kidney	4.6	Pituitary gland Pool	0.8
Renal ca. 786-0	17.2	Salivary Gland	5.6
Renal ca. A498	5.1	Thyroid (female)	9.7
Renal ca. ACHN	17.3	Pancreatic ca. CAPAN2	11.7
Renal ca. UO-31	11.1	Pancreas Pool	9.2

CNS_neurodegeneration_v1.0 Summary: Ag3362 Highest expression of the CG58516-01 gene is seen in the occipital cortex of a control patient and the temporal cortex of an Alzheimer's patient. While the CG58516-01 gene does not appear to be preferentially expressed in Alzheimer's disease, this panel confirms expression of the CG58516-01 gene at moderate/high levels in the brain in an additional set of individuals. Please see Panel 1.4 for discussion of potential utility of this gene in the central nervous system.

General_screening_panel_v1.4 Summary: Ag3362 The CG58516-01 gene is widely expressed in this panel, with highest expression in the breast cancer cell line T47D (CT=29). Significant expression is also seen in cell lines derived from prostate, breast and ovarian cancers. In general, expression of the CG58516-01 gene appears to be greater in the cancer cell lines than in normal tissue. Thus, the expression of this gene could be used to distinguish these cell line types from others in the panel.

Among tissues involved in central nervous system function, this gene is expressed at low but significant levels in all brain regions examined. This gene encodes a protein with a putative zinc-finger motif. Since these proteins are known to interact with nucleic acids,

515

this suggests that this gene product may play a potential role in transcription. Thus, therapeutic modulation of the CG58516-01 gene product may be used to regulate the transcription of disease-related proteins such as ataxin, huntingtin, or various apoptosis cascade proteins.

5   Among tissues with metabolic function, this gene is expressed at low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, skeletal muscle, heart, and fetal liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic and that disregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and

10   diabetes.

References:

        1. Zhu W, Chan EK, Li J, Hemmerich P, Tan EM. (2001) Transcription activating property of autoantigen SG2NA and modulating effect of WD-40 repeats. Exp Cell Res. 269(2):312-21

15   **Panel 4D Summary:** Ag3362 Results from one experiment with the CG58516-01 gene are not included because the amp plot corresponding to the run indicates that there were problems with the experiment.

F. CG58473-01: PROTEIN KINASE

        Expression of gene CG58473-01 was assessed using the primer-probe set Ag3357,

20   described in Table FA. Results of the RTQ-PCR runs are shown in Tables FB and FC.

Table FA. Probe Name Ag3357

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtcaaggtggccctaaaattc-3'	21	853	384
Probe	TET-5'-ccaggacctcatctccaagctgctta-3'-TAMRA	26	897	385
Reverse	5'-agccgttctgaggggttat-3'	19	926	386

Table FB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%)	Tissue Name	Rel. Exp.(%)

	Ag3357, Run 216523477		Ag3357, Run 216523477
Adipose	0.0	Renal ca. TK-10	13.2
Melanoma* Hs688(A).T	0.0	Bladder	7.2
Melanoma* Hs688(B).T	1.1	Gastric ca. (liver met.) NCI-N87	5.4
Melanoma* M14	50.0	Gastric ca. KATO III	49.0
Melanoma* LOXIMVI	33.0	Colon ca. SW-948	14.9
Melanoma* SK-MEL-5	24.7	Colon ca. SW480	95.9
Squamous cell carcinoma SCC-4	11.6	Colon ca.* (SW480 met) SW620	53.6
Testis Pool	8.2	Colon ca. HT29	10.3
Prostate ca.* (bone met) PC-3	3.2	Colon ca. HCT-116	76.3
Prostate Pool	0.0	Colon ca. CaCo-2	14.1
Placenta	2.4	Colon cancer tissue	6.3
Uterus Pool	0.0	Colon ca. SW1116	18.6
Ovarian ca. OVCAR-3	51.1	Colon ca. Colo-205	24.3
Ovarian ca. SK-OV-3	53.2	Colon ca. SW-48	26.1
Ovarian ca. OVCAR-4	10.4	Colon Pool	4.6
Ovarian ca. OVCAR-5	12.3	Small Intestine Pool	1.7
Ovarian ca. IGROV-1	10.1	Stomach Pool	1.2
Ovarian ca. OVCAR-8	13.4	Bone Marrow Pool	1.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	20.3	Heart Pool	0.0
Breast ca. MDA-MB-231	65.1	Lymph Node Pool	1.4
Breast ca. BT 549	**100.0**	Fetal Skeletal Muscle	0.0
Breast ca. T47D	34.2	Skeletal Muscle Pool	1.6
Breast ca. MDA-N	36.3	Spleen Pool	3.4
Breast Pool	1.3	Thymus Pool	4.7
Trachea	0.0	CNS cancer (glio/astro) U87-MG	7.8
Lung	0.0	CNS cancer (glio/astro) U-118-MG	54.0

Fetal Lung	5.0	CNS cancer (neuro;met) SK-N-AS	7.9
Lung ca. NCI-N417	17.9	CNS cancer (astro) SF-539	22.4
Lung ca. LX-1	28.5	CNS cancer (astro) SNB-75	19.2
Lung ca. NCI-H146	74.7	CNS cancer (glio) SNB-19	14.6
Lung ca. SHP-77	40.6	CNS cancer (glio) SF-295	3.0
Lung ca. A549	64.6	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	23.8	Brain (cerebellum)	0.0
Lung ca. NCI-H23	63.7	Brain (fetal)	0.0
Lung ca. NCI-H460	0.8	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	2.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	34.4	Brain (Substantia nigra) Pool	2.6
Liver	0.0	Brain (Thalamus) Pool	9.3
Fetal Liver	0.0	Brain (whole)	2.5
Liver ca. HepG2	11.4	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	3.1	Pituitary gland Pool	1.4
Renal ca. 786-0	20.0	Salivary Gland	0.0
Renal ca. A498	3.6	Thyroid (female)	0.0
Renal ca. ACHN	18.9	Pancreatic ca. CAPAN2	20.4
Renal ca. UO-31	10.4	Pancreas Pool	1.3

Table FC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3357, Run 165231196	Tissue Name	Rel. Exp.(%) Ag3357, Run 165231196
Secondary Th1 act	9.0	HUVEC IL-1beta	9.5
Secondary Th2 act	43.2	HUVEC IFN gamma	6.3
Secondary Tr1 act	46.0	HUVEC TNF alpha + IFN gamma	7.3
Secondary Th1 rest	6.7	HUVEC TNF alpha + IL4	25.3
Secondary Th2 rest	12.2	HUVEC IL-11	13.1
Secondary Tr1 rest	1.9	Lung Microvascular EC none	3.3
Primary Th1 act	6.1	Lung Microvascular EC	7.1

		TNFalpha + IL-1beta	
Primary Th2 act	21.8	Microvascular Dermal EC none	10.9
Primary Tr1 act	33.0	Microsvasular Dermal EC TNFalpha + IL-1beta	7.3
Primary Th1 rest	28.1	Bronchial epithelium TNFalpha + IL1beta	1.9
Primary Th2 rest	12.1	Small airway epithelium none	3.6
Primary Tr1 rest	29.7	Small airway epithelium TNFalpha + IL-1beta	36.3
CD45RA CD4 lymphocyte act	28.5	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	39.8	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	18.6	Astrocytes rest	1.4
Secondary CD8 lymphocyte rest	26.8	Astrocytes TNFalpha + IL-1beta	1.2
Secondary CD8 lymphocyte act	19.2	KU-812 (Basophil) rest	18.2
CD4 lymphocyte none	10.6	KU-812 (Basophil) PMA/ionomycin	30.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	15.6	CCD1106 (Keratinocytes) none	18.3
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	7.7
LAK cells IL-2	42.6	Liver cirrhosis	25.7
LAK cells IL-2+IL-12	24.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	24.8	NCI-H292 none	7.8
LAK cells IL-2+ IL-18	40.3	NCI-H292 IL-4	26.4
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	29.7
NK Cells IL-2 rest	23.5	NCI-H292 IL-13	20.7
Two Way MLR 3 day	13.7	NCI-H292 IFN gamma	27.9
Two Way MLR 5 day	13.2	HPAEC none	8.6
Two Way MLR 7 day	11.7	HPAEC TNF alpha + IL-1 beta	2.4
PBMC rest	0.0	Lung fibroblast none	5.5
PBMC PWM	52.1	Lung fibroblast TNF alpha + IL-1 beta	2.2
PBMC PHA-L	14.6	Lung fibroblast IL-4	0.0
Ramos (B cell) none	16.5	Lung fibroblast IL-9	0.0
Ramos (B cell)	14.7	Lung fibroblast IL-13	0.0

ionomycin			
B lymphocytes PWM	**100.0**	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	10.4	Dermal fibroblast CCD1070 rest	40.1
EOL-1 dbcAMP	9.9	Dermal fibroblast CCD1070 TNF alpha	43.8
EOL-1 dbcAMP PMA/ionomycin	13.2	Dermal fibroblast CCD1070 IL-1 beta	23.5
Dendritic cells none	4.7	Dermal fibroblast IFN gamma	3.7
Dendritic cells LPS	1.1	Dermal fibroblast IL-4	4.6
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	28.1
Macrophages rest	4.3	Lung	59.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	28.3	Kidney	10.0
HUVEC starved	25.3		

CNS_neurodegeneration_v1.0 Summary: Ag3357 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3357 This gene is primarily expressed in cancer cell lines, with highest expression in a breast cancer cell line BT 549(CT=32.8). This gene is expressed in the following cell lines but not the corresponding healthy tissue: gastric, brain, colon, lung, breast, ovarian cancer and melanomas. Thus, expression of this gene could be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition using antibodies or small molecule drugs might be of use in the treatment of these cancers.

Panel 4D Summary: Ag3357 Highest expression of the CG58473-01 gene is seen in pokeweed mitogen-activated purified peripheral blood B lymphocytes (CT=33.2). In addition, no expression of the transcript is seen in PBMC that contain normal B cells, but the transcript is induced when PBMC are treated with the B cell selective pokeweed mitogen. The transcript is not seen in the B cell lymphoma cell line Ramos regardless of stimulation. Thus, the putative protein encoded by this gene could potentially be used diagnostically to identify activated B cells. Therefore, therapeutics that antagonize the function of this gene product may be useful as therapeutic drugs to reduce or eliminate the

symptoms in patients with autoimmune and inflammatory diseases in which B cells play a part in the intiation or progression of the disease process, such as lupus erythematosus, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, or psoriasis.

G. CG58470-01: UDP-N-ACETYLHEXOSAMINE PYROPHOSPHORYLASE

Expression of gene CG58470-01 was assessed using the primer-probe set Ag5940, described in Table GA.

Table GA. Probe Name Ag5940

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atatcctgaagctacaacagttagct-3'	26	422	387
Probe	TET-5'-tggcaacaaatgcattattccatattacg-3'-TAMRA	29	459	388
Reverse	5'-gagtgaactcgctggtcatg-3'	20	489	389

General_screening_panel_v1.5 Summary: Ag5940 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 5 Islet Summary: Ag5940 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

H. CG58593-01: UBIQUITIN-52

Expression of gene CG58593-01 was assessed using the primer-probe set Ag3421, described in Table HA.

Table HA. Probe Name Ag3421

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atctgctgcaagtgctatgc-3'	20	291	390
Probe	TET-5'-cggtgctatcaactgccacaagaaga-3'-TAMRA	26	323	391
Reverse	5'-tgaccttcttcctggggtac-3'	20	371	392

CNS_neurodegeneration_v1.0 Summary: Ag3421 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3421 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

5 **Panel 4D Summary:** Ag3421 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

I. CG57871-01: TOUSLED-LIKE KINASE

Expression of gene CG57871-01 was assessed using the primer-probe set Ag3351, described in Table IA. Results of the RTQ-PCR runs are shown in Tables IB and IC.

10 Table IA. Probe Name Ag3351

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gatcctcactgcaacattcttt-3'	22	346	393
Probe	TET-5'-aatcccttaccgcgacgagtagaaca-3'-TAMRA	26	372	394
Reverse	5'-gcactgccatctaaaccataga-3'	22	403	395

Table IB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3351, Run 210141594	Tissue Name	Rel. Exp.(%) Ag3351, Run 210141594
AD 1 Hippo	10.4	Control (Path) 3 Temporal Ctx	3.0
AD 2 Hippo	33.4	Control (Path) 4 Temporal Ctx	65.1
AD 3 Hippo	5.5	AD 1 Occipital Ctx	20.2
AD 4 Hippo	8.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	**100.0**	AD 3 Occipital Ctx	3.8
AD 6 Hippo	33.4	AD 4 Occipital Ctx	45.1
Control 2 Hippo	29.9	AD 5 Occipital Ctx	15.2
Control 4 Hippo	6.7	AD 6 Occipital Ctx	46.7

Control (Path) 3 Hippo	3.7	Control 1 Occipital Ctx	2.7
AD 1 Temporal Ctx	16.8	Control 2 Occipital Ctx	52.5
AD 2 Temporal Ctx	45.1	Control 3 Occipital Ctx	45.4
AD 3 Temporal Ctx	6.9	Control 4 Occipital Ctx	6.3
AD 4 Temporal Ctx	54.0	Control (Path) 1 Occipital Ctx	79.0
AD 5 Inf Temporal Ctx	92.0	Control (Path) 2 Occipital Ctx	34.4
AD 5 SupTemporal Ctx	13.0	Control (Path) 3 Occipital Ctx	0.8
AD 6 Inf Temporal Ctx	48.6	Control (Path) 4 Occipital Ctx	40.6
AD 6 Sup Temporal Ctx	56.6	Control 1 Parietal Ctx	6.9
Control 1 Temporal Ctx	6.2	Control 2 Parietal Ctx	48.0
Control 2 Temporal Ctx	29.3	Control 3 Parietal Ctx	26.1
Control 3 Temporal Ctx	32.8	Control (Path) 1 Parietal Ctx	73.7
Control 4 Temporal Ctx	13.9	Control (Path) 2 Parietal Ctx	57.4
Control (Path) 1 Temporal Ctx	79.6	Control (Path) 3 Parietal Ctx	3.4
Control (Path) 2 Temporal Ctx	97.3	Control (Path) 4 Parietal Ctx	78.5

Table IC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3351, Run 165222896	Tissue Name	Rel. Exp.(%) Ag3351, Run 165222896
Secondary Th1 act	16.5	HUVEC IL-1beta	15.4
Secondary Th2 act	26.4	HUVEC IFN gamma	13.5
Secondary Tr1 act	23.3	HUVEC TNF alpha + IFN gamma	17.0
Secondary Th1 rest	6.0	HUVEC TNF alpha + IL4	11.0
Secondary Th2 rest	10.7	HUVEC IL-11	5.4
Secondary Tr1 rest	2.1	Lung Microvascular EC none	12.4

Primary Th1 act	19.2	Lung Microvascular EC TNFalpha + IL-1beta	9.6
Primary Th2 act	17.6	Microvascular Dermal EC none	14.7
Primary Tr1 act	36.1	Microsvasular Dermal EC TNFalpha + IL-1beta	14.8
Primary Th1 rest	55.5	Bronchial epithelium TNFalpha + IL1beta	14.1
Primary Th2 rest	43.8	Small airway epithelium none	7.7
Primary Tr1 rest	15.9	Small airway epithelium TNFalpha + IL-1beta	50.3
CD45RA CD4 lymphocyte act	13.0	Coronery artery SMC rest	15.6
CD45RO CD4 lymphocyte act	21.0	Coronery artery SMC TNFalpha + IL-1beta	6.1
CD8 lymphocyte act	12.9	Astrocytes rest	11.5
Secondary CD8 lymphocyte rest	14.9	Astrocytes TNFalpha + IL-1beta	11.8
Secondary CD8 lymphocyte act	14.8	KU-812 (Basophil) rest	19.2
CD4 lymphocyte none	10.7	KU-812 (Basophil) PMA/ionomycin	54.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	12.7	CCD1106 (Keratinocytes) none	12.2
LAK cells rest	17.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	9.0
LAK cells IL-2	22.4	Liver cirrhosis	7.4
LAK cells IL-2+IL-12	20.4	Lupus kidney	3.4
LAK cells IL-2+IFN gamma	37.9	NCI-H292 none	47.6
LAK cells IL-2+ IL-18	18.6	NCI-H292 IL-4	42.3
LAK cells PMA/ionomycin	10.5	NCI-H292 IL-9	30.4
NK Cells IL-2 rest	17.8	NCI-H292 IL-13	15.7
Two Way MLR 3 day	33.2	NCI-H292 IFN gamma	25.5
Two Way MLR 5 day	10.6	HPAEC none	13.5
Two Way MLR 7 day	9.9	HPAEC TNF alpha + IL-1 beta	17.7
PBMC rest	12.8	Lung fibroblast none	11.5
PBMC PWM	63.3	Lung fibroblast TNF alpha + IL-1 beta	12.4
PBMC PHA-L	18.0	Lung fibroblast IL-4	31.2
Ramos (B cell) none	14.0	Lung fibroblast IL-9	22.2

Ramos (B cell) ionomycin	77.9	Lung fibroblast IL-13	27.4
B lymphocytes PWM	**100.0**	Lung fibroblast IFN gamma	44.8
B lymphocytes CD40L and IL-4	30.8	Dermal fibroblast CCD1070 rest	33.7
EOL-1 dbcAMP	11.3	Dermal fibroblast CCD1070 TNF alpha	50.0
EOL-1 dbcAMP PMA/ionomycin	13.7	Dermal fibroblast CCD1070 IL-1 beta	13.4
Dendritic cells none	14.7	Dermal fibroblast IFN gamma	14.3
Dendritic cells LPS	19.8	Dermal fibroblast IL-4	25.7
Dendritic cells anti-CD40	14.2	IBD Colitis 2	2.0
Monocytes rest	22.5	IBD Crohn's	3.2
Monocytes LPS	32.8	Colon	26.8
Macrophages rest	31.0	Lung	14.6
Macrophages LPS	30.8	Thymus	28.7
HUVEC none	18.3	Kidney	45.4
HUVEC starved	45.7		

CNS_neurodegeneration_v1.0 Summary: Ag3351 - This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. While no differential expression of this gene is detected between Alzheimer's diseased postmortem brains and those of non-demented controls, the widespread expression of this gene in the brain suggests that therapeutic modulation of the expression or function of this gene may be effective in the treatment of neurologic disorders such as Parkinson's disease, epilepsy, stroke and multiple sclerosis.

General_screening_panel_v1.4 Summary: Ag3351 - Results from one experiment are not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 4D Summary: Ag3351 The CG57871-01 gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This

pattern also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

J. CG58590-01 and CG58590-02: PALS Guanylate kinase

Expression of gene CG58590-01 and CG58590-02 was assessed using the primer-probe set Ag3380, described in Table JA. Results of the RTQ-PCR runs are shown in Tables JB, JC and JD. Please note that CG58590-02 represents a full-length physical clone of the CG58590-01 gene, validating the prediction of the gene sequence.

Table JA. Probe Name Ag3380

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tttgatacggcaattgtgaatt-3'	22	1931	396
Probe	TET-5'-ccgatcttgataaagcctatcaggaa-3'-TAMRA	26	1953	397
Reverse	5'-cccactgaggttcagtatcaag-3'	22	2000	398

Table JB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3380, Run 210153753	Tissue Name	Rel. Exp.(%) Ag3380, Run 210153753
AD 1 Hippo	12.9	Control (Path) 3 Temporal Ctx	4.7
AD 2 Hippo	27.7	Control (Path) 4 Temporal Ctx	24.3
AD 3 Hippo	4.8	AD 1 Occipital Ctx	15.6
AD 4 Hippo	7.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	**100.0**	AD 3 Occipital Ctx	7.5
AD 6 Hippo	64.2	AD 4 Occipital Ctx	19.1
Control 2 Hippo	25.5	AD 5 Occipital Ctx	29.5
Control 4 Hippo	9.9	AD 6 Occipital	40.1

526

		Ctx	
Control (Path) 3 Hippo	8.4	Control 1 Occipital Ctx	4.2
AD 1 Temporal Ctx	17.6	Control 2 Occipital Ctx	65.5
AD 2 Temporal Ctx	25.3	Control 3 Occipital Ctx	13.4
AD 3 Temporal Ctx	4.9	Control 4 Occipital Ctx	6.4
AD 4 Temporal Ctx	17.4	Control (Path) 1 Occipital Ctx	78.5
AD 5 Inf Temporal Ctx	81.8	Control (Path) 2 Occipital Ctx	9.4
AD 5 SupTemporal Ctx	42.9	Control (Path) 3 Occipital Ctx	3.2
AD 6 Inf Temporal Ctx	48.6	Control (Path) 4 Occipital Ctx	9.9
AD 6 Sup Temporal Ctx	53.6	Control 1 Parietal Ctx	6.0
Control 1 Temporal Ctx	5.7	Control 2 Parietal Ctx	37.1
Control 2 Temporal Ctx	34.6	Control 3 Parietal Ctx	16.5
Control 3 Temporal Ctx	10.2	Control (Path) 1 Parietal Ctx	67.4
Control 4 Temporal Ctx	7.1	Control (Path) 2 Parietal Ctx	18.7
Control (Path) 1 Temporal Ctx	41.5	Control (Path) 3 Parietal Ctx	3.3
Control (Path) 2 Temporal Ctx	29.5	Control (Path) 4 Parietal Ctx	34.4

Table JC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3380, Run 217043276	Tissue Name	Rel. Exp.(%) Ag3380, Run 217043276
Adipose	9.0	Renal ca. TK-10	25.5
Melanoma* Hs688(A).T	18.9	Bladder	15.9
Melanoma* Hs688(B).T	16.8	Gastric ca. (liver met.) NCI-N87	52.5
Melanoma* M14	14.9	Gastric ca. KATO III	34.6
Melanoma* LOXIMVI	21.6	Colon ca. SW-948	4.9

527

Melanoma* SK-MEL-5	27.0	Colon ca. SW480	82.4
Squamous cell carcinoma SCC-4	28.7	Colon ca.* (SW480 met) SW620	20.6
Testis Pool	5.1	Colon ca. HT29	9.2
Prostate ca.* (bone met) PC-3	59.9	Colon ca. HCT-116	20.6
Prostate Pool	8.6	Colon ca. CaCo-2	22.8
Placenta	3.9	Colon cancer tissue	10.1
Uterus Pool	1.9	Colon ca. SW1116	6.2
Ovarian ca. OVCAR-3	32.5	Colon ca. Colo-205	4.9
Ovarian ca. SK-OV-3	57.4	Colon ca. SW-48	4.2
Ovarian ca. OVCAR-4	14.7	Colon Pool	11.4
Ovarian ca. OVCAR-5	59.5	Small Intestine Pool	9.8
Ovarian ca. IGROV-1	13.1	Stomach Pool	7.4
Ovarian ca. OVCAR-8	19.2	Bone Marrow Pool	4.2
Ovary	5.9	Fetal Heart	6.3
Breast ca. MCF-7	35.1	Heart Pool	4.9
Breast ca. MDA-MB-231	58.2	Lymph Node Pool	11.4
Breast ca. BT 549	26.8	Fetal Skeletal Muscle	3.3
Breast ca. T47D	**100.0**	Skeletal Muscle Pool	8.1
Breast ca. MDA-N	8.7	Spleen Pool	5.6
Breast Pool	10.4	Thymus Pool	6.3
Trachea	5.5	CNS cancer (glio/astro) U87-MG	39.2
Lung	3.8	CNS cancer (glio/astro) U-118-MG	54.7
Fetal Lung	11.8	CNS cancer (neuro;met) SK-N-AS	19.6
Lung ca. NCI-N417	3.2	CNS cancer (astro) SF-539	12.2
Lung ca. LX-1	20.7	CNS cancer (astro) SNB-75	29.7
Lung ca. NCI-H146	3.8	CNS cancer (glio) SNB-19	13.4
Lung ca. SHP-77	17.9	CNS cancer (glio) SF-295	28.9
Lung ca. A549	30.6	Brain (Amygdala) Pool	11.8

Lung ca. NCI-H526	3.6	Brain (cerebellum)	6.0
Lung ca. NCI-H23	29.3	Brain (fetal)	8.4
Lung ca. NCI-H460	14.8	Brain (Hippocampus) Pool	14.5
Lung ca. HOP-62	19.5	Cerebral Cortex Pool	16.2
Lung ca. NCI-H522	28.7	Brain (Substantia nigra) Pool	16.0
Liver	0.4	Brain (Thalamus) Pool	22.7
Fetal Liver	11.9	Brain (whole)	5.9
Liver ca. HepG2	12.9	Spinal Cord Pool	16.0
Kidney Pool	18.4	Adrenal Gland	5.1
Fetal Kidney	22.8	Pituitary gland Pool	3.8
Renal ca. 786-0	28.5	Salivary Gland	2.1
Renal ca. A498	5.0	Thyroid (female)	8.2
Renal ca. ACHN	22.4	Pancreatic ca. CAPAN2	51.4
Renal ca. UO-31	36.9	Pancreas Pool	12.3

Table JD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3380, Run 165296532	Tissue Name	Rel. Exp.(%) Ag3380, Run 165296532
Secondary Th1 act	13.1	HUVEC IL-1beta	15.0
Secondary Th2 act	14.6	HUVEC IFN gamma	19.6
Secondary Tr1 act	15.2	HUVEC TNF alpha + IFN gamma	28.3
Secondary Th1 rest	4.6	HUVEC TNF alpha + IL4	26.1
Secondary Th2 rest	4.7	HUVEC IL-11	7.8
Secondary Tr1 rest	8.0	Lung Microvascular EC none	25.5
Primary Th1 act	14.9	Lung Microvascular EC TNFalpha + IL-1beta	19.5
Primary Th2 act	13.2	Microvascular Dermal EC none	37.9
Primary Tr1 act	20.7	Microsvasular Dermal EC TNFalpha + IL-1beta	24.8
Primary Th1 rest	35.6	Bronchial epithelium TNFalpha + IL1beta	37.1
Primary Th2 rest	24.0	Small airway epithelium none	15.0
Primary Tr1 rest	16.2	Small airway epithelium TNFalpha + IL-1beta	**100.0**

CD45RA CD4 lymphocyte act	23.3	Coronery artery SMC rest	30.1
CD45RO CD4 lymphocyte act	18.2	Coronery artery SMC TNFalpha + IL-1beta	13.6
CD8 lymphocyte act	7.4	Astrocytes rest	22.5
Secondary CD8 lymphocyte rest	13.4	Astrocytes TNFalpha + IL-1beta	21.2
Secondary CD8 lymphocyte act	4.4	KU-812 (Basophil) rest	17.9
CD4 lymphocyte none	8.0	KU-812 (Basophil) PMA/ionomycin	68.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	10.7	CCD1106 (Keratinocytes) none	22.1
LAK cells rest	13.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	9.2
LAK cells IL-2	12.9	Liver cirrhosis	3.1
LAK cells IL-2+IL-12	13.2	Lupus kidney	2.9
LAK cells IL-2+IFN gamma	15.6	NCI-H292 none	48.6
LAK cells IL-2+ IL-18	17.0	NCI-H292 IL-4	66.9
LAK cells PMA/ionomycin	9.5	NCI-H292 IL-9	59.5
NK Cells IL-2 rest	7.0	NCI-H292 IL-13	36.6
Two Way MLR 3 day	15.2	NCI-H292 IFN gamma	42.6
Two Way MLR 5 day	7.0	HPAEC none	14.3
Two Way MLR 7 day	9.6	HPAEC TNF alpha + IL-1 beta	25.9
PBMC rest	6.4	Lung fibroblast none	12.5
PBMC PWM	60.7	Lung fibroblast TNF alpha + IL-1 beta	11.0
PBMC PHA-L	18.8	Lung fibroblast IL-4	25.9
Ramos (B cell) none	31.9	Lung fibroblast IL-9	20.6
Ramos (B cell) ionomycin	94.0	Lung fibroblast IL-13	18.8
B lymphocytes PWM	42.9	Lung fibroblast IFN gamma	23.3
B lymphocytes CD40L and IL-4	24.7	Dermal fibroblast CCD1070 rest	59.5
EOL-1 dbcAMP	12.9	Dermal fibroblast CCD1070 TNF alpha	64.2
EOL-1 dbcAMP PMA/ionomycin	10.4	Dermal fibroblast CCD1070 IL-1 beta	32.8
Dendritic cells none	19.6	Dermal fibroblast IFN gamma	10.7

Dendritic cells LPS	10.7	Dermal fibroblast IL-4	21.6
Dendritic cells anti-CD40	18.8	IBD Colitis 2	2.0
Monocytes rest	15.0	IBD Crohn's	3.6
Monocytes LPS	13.8	Colon	36.9
Macrophages rest	25.3	Lung	19.3
Macrophages LPS	8.1	Thymus	72.2
HUVEC none	19.9	Kidney	24.5
HUVEC starved	35.8		

CNS_neurodegeneration_v1.0 Summary: Ag3380 This panel does not show differential expression of the CG58590-01 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

5   General_screening_panel_v1.4 Summary: Ag3380 - This gene is expressed at low to moderate levels in all samples on this pattern. The highest level of expression is seen in breast cancer cell line T47D (CT=27.8). Based on expression in this panel, this gene may be involved in brain, colon, renal, lung, ovarian and prostate cancer as well as melanomas. Thus, expression of this gene could be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition using antibodies or small molecule drugs might be of use in the treatment of these cancers.

This gene product is also expressed in adipose, pancreas, adrenal, thyroid, pituitary, skeletal muscle, heart, and fetal liver. This widespread expression in tissues with metabolic function suggests that this gene product may be important for the pathogenesis, diagnosis, and/or treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes. Furthermore, this gene is more highly expressed in fetal (CT=30.9) liver when compared to expression in the adult (CT>35) and may be useful for the differentiation of the fetal and adult sources of this tissue.

In addition, this gene is expressed at moderate levels in the all regions of the CNS examined. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

531

Panel 4D Summary: Ag3380 - This gene is expressed from moderate to low levels across all of the samples on this panel. The highest expression is seen in small airway epithelium treated with TNFalpha and IL-1beta (CT=28.7). Interestingly, expression is much lower in untreated small airway epithelium (CT=31.5). There is also a significant difference between mononuclear cells treated with PWM (CT=29.5) and untreated cells (CT=32.7). Therefore, expression of this gene can be used to differentiate treated and untreated samples.

Expression of this gene is detected at a moderate level (CT=30.2) in normal colon (similar levels for colon are seen on panel 1.4 (CT=30.9), but is significantly lower in the IBD Colitis 2 (CT=34.4) and IBD Crohn's (CT=33.5) samples. Therefore, therapies designed with the protein encoded for by this gene may potentially modulate colon function and play a role in the identification and treatment of inflammatory or autoimmune diseases which effect the colon including Crohn's disease and ulcerative colitis.

K. CG58572-01 and CG58572-02: GLUCOSAMINE-PHOSPHATE N-ACETYLTRANSFERASE

Expression of gene CG58572-01 and full length clone CG58572-02 was assessed using the primer-probe set Ag3375, described in Table KA. Results of the RTQ-PCR runs are shown in Tables KB, KC and KD.

Table KA. Probe Name Ag3375

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aagaagtggactggagtcagaa-3'	22	58	399
Probe	TET-5'-tacattttctccagccatttccccaa-3'-TAMRA	26	86	400
Reverse	5'-agcagtacaaagaggcctcaa-3'	21	135	401

Table KB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3375, Run 210154239	Tissue Name	Rel. Exp.(%) Ag3375, Run 210154239
AD 1 Hippo	17.1	Control (Path) 3 Temporal Ctx	4.8
AD 2 Hippo	19.3	Control (Path) 4 Temporal Ctx	27.5
AD 3 Hippo	7.4	AD 1 Occipital Ctx	11.5

AD 4 Hippo	4.5	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	72.2	AD 3 Occipital Ctx	5.9
AD 6 Hippo	53.6	AD 4 Occipital Ctx	12.7
Control 2 Hippo	20.3	AD 5 Occipital Ctx	26.6
Control 4 Hippo	6.8	AD 6 Occipital Ctx	19.8
Control (Path) 3 Hippo	5.5	Control 1 Occipital Ctx	3.2
AD 1 Temporal Ctx	11.6	Control 2 Occipital Ctx	36.1
AD 2 Temporal Ctx	23.8	Control 3 Occipital Ctx	7.4
AD 3 Temporal Ctx	5.5	Control 4 Occipital Ctx	4.1
AD 4 Temporal Ctx	16.5	Control (Path) 1 Occipital Ctx	66.0
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	8.2
AD 5 Sup Temporal Ctx	55.9	Control (Path) 3 Occipital Ctx	1.9
AD 6 Inf Temporal Ctx	37.9	Control (Path) 4 Occipital Ctx	12.2
AD 6 Sup Temporal Ctx	59.5	Control 1 Parietal Ctx	2.4
Control 1 Temporal Ctx	3.5	Control 2 Parietal Ctx	31.6
Control 2 Temporal Ctx	25.3	Control 3 Parietal Ctx	11.7
Control 3 Temporal Ctx	8.2	Control (Path) 1 Parietal Ctx	49.7
Control 3 Temporal Ctx	4.0	Control (Path) 2 Parietal Ctx	15.4
Control (Path) 1 Temporal Ctx	52.9	Control (Path) 3 Parietal Ctx	4.2
Control (Path) 2 Temporal Ctx	26.6	Control (Path) 4 Parietal Ctx	32.5

Table KC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3375, Run 165674233	Tissue Name	Rel. Exp.(%) Ag3375, Run 165674233
Liver adenocarcinoma	51.8	Kidney (fetal)	9.7
Pancreas	9.3	Renal ca. 786-0	19.6
Pancreatic ca. CAPAN	52.1	Renal ca. A498	26.2

2			
Adrenal gland	8.9	Renal ca. RXF 393	15.7
Thyroid	6.3	Renal ca. ACHN	8.2
Salivary gland	18.3	Renal ca. UO-31	35.4
Pituitary gland	15.1	Renal ca. TK-10	9.8
Brain (fetal)	15.5	Liver	20.4
Brain (whole)	34.6	Liver (fetal)	16.5
Brain (amygdala)	16.0	Liver ca. (hepatoblast) HepG2	49.0
Brain (cerebellum)	34.2	Lung	4.5
Brain (hippocampus)	12.1	Lung (fetal)	5.4
Brain (substantia nigra)	12.8	Lung ca. (small cell) LX-1	32.3
Brain (thalamus)	17.9	Lung ca. (small cell) NCI-H69	17.3
Cerebral Cortex	10.4	Lung ca. (s.cell var.) SHP-77	30.1
Spinal cord	13.3	Lung ca. (large cell)NCI-H460	66.4
glio/astro U87-MG	14.8	Lung ca. (non-sm. cell) A549	19.1
glio/astro U-118-MG	95.3	Lung ca. (non-s.cell) NCI-H23	13.8
Astrocytoma SW1783	42.0	Lung ca. (non-s.cell) HOP-62	18.7
neuro*; met SK-N-AS	47.0	Lung ca. (non-s.cl) NCI-H522	19.5
Astrocytoma SF-539	11.4	Lung ca. (squam.) SW 900	9.9
Astrocytoma SNB-75	15.6	Lung ca. (squam.) NCI-H596	19.6
glioma SNB-19	11.8	Mammary gland	14.6
glioma U251	40.9	Breast ca.* (pl.ef) MCF-7	81.2
glioma SF-295	10.1	Breast ca.* (pl.ef) MDA-MB-231	91.4
Heart (fetal)	1.3	Breast ca.* (pl.ef) T47D	35.4
Heart	4.7	Breast ca. BT-549	97.9
Skeletal muscle (fetal)	1.2	Breast ca. MDA-N	14.8
Skeletal muscle	38.7	Ovary	1.6
Bone marrow	4.6	Ovarian ca. OVCAR-3	39.2
Thymus	2.7	Ovarian ca.	23.0

534

		OVCAR-4	
Spleen	7.9	Ovarian ca. OVCAR-5	13.8
Lymph node	13.0	Ovarian ca. OVCAR-8	8.5
Colorectal	3.3	Ovarian ca. IGROV-1	5.6
Stomach	27.7	Ovarian ca.* (ascites) SK-OV-3	44.8
Small intestine	19.3	Uterus	19.5
Colon ca. SW480	16.5	Placenta	2.6
Colon ca.* SW620(SW480 met)	29.1	Prostate	15.6
Colon ca. HT29	13.8	Prostate ca.* (bone met)PC-3	56.6
Colon ca. HCT-116	27.7	Testis	40.6
Colon ca. CaCo-2	17.4	Melanoma Hs688(A).T	5.5
Colon ca. tissue(ODO3866)	26.4	Melanoma* (met) Hs688(B).T	8.9
Colon ca. HCC-2998	32.1	Melanoma UACC-62	17.8
Gastric ca.* (liver met) NCI-N87	**100.0**	Melanoma M14	27.7
Bladder	28.7	Melanoma LOX IMVI	6.6
Trachea	9.4	Melanoma* (met) SK-MEL-5	13.0
Kidney	9.0	Adipose	8.0

Table KD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3375, Run 165296547	Tissue Name	Rel. Exp.(%) Ag3375, Run 165296547
Secondary Th1 act	14.6	HUVEC IL-1beta	24.5
Secondary Th2 act	13.0	HUVEC IFN gamma	24.5
Secondary Tr1 act	17.3	HUVEC TNF alpha + IFN gamma	24.0
Secondary Th1 rest	0.9	HUVEC TNF alpha + IL4	23.2
Secondary Th2 rest	1.5	HUVEC IL-11	12.1
Secondary Tr1 rest	2.9	Lung Microvascular EC none	21.3
Primary Th1 act	16.0	Lung Microvascular EC	24.1

535

		TNFalpha + IL-1beta	
Primary Th2 act	12.1	Microvascular Dermal EC none	27.4
Primary Tr1 act	25.0	Microsvasular Dermal EC TNFalpha + IL-1beta	24.0
Primary Th1 rest	10.4	Bronchial epithelium TNFalpha + IL1beta	20.3
Primary Th2 rest	6.1	Small airway epithelium none	11.3
Primary Tr1 rest	9.0	Small airway epithelium TNFalpha + IL-1beta	54.0
CD45RA CD4 lymphocyte act	14.6	Coronery artery SMC rest	23.5
CD45RO CD4 lymphocyte act	13.6	Coronery artery SMC TNFalpha + IL-1beta	12.0
CD8 lymphocyte act	14.2	Astrocytes rest	5.3
Secondary CD8 lymphocyte rest	14.4	Astrocytes TNFalpha + IL-1beta	5.4
Secondary CD8 lymphocyte act	5.8	KU-812 (Basophil) rest	19.5
CD4 lymphocyte none	2.4	KU-812 (Basophil) PMA/ionomycin	56.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	2.6	CCD1106 (Keratinocytes) none	26.6
LAK cells rest	5.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	7.8
LAK cells IL-2	10.7	Liver cirrhosis	2.6
LAK cells IL-2+IL-12	12.5	Lupus kidney	0.8
LAK cells IL-2+IFN gamma	20.2	NCI-H292 none	28.7
LAK cells IL-2+ IL-18	16.6	NCI-H292 IL-4	54.7
LAK cells PMA/ionomycin	12.5	NCI-H292 IL-9	45.7
NK Cells IL-2 rest	7.1	NCI-H292 IL-13	24.3
Two Way MLR 3 day	6.8	NCI-H292 IFN gamma	33.2
Two Way MLR 5 day	8.9	HPAEC none	17.8
Two Way MLR 7 day	6.0	HPAEC TNF alpha + IL-1 beta	30.1
PBMC rest	0.8	Lung fibroblast none	10.2
PBMC PWM	42.3	Lung fibroblast TNF alpha + IL-1 beta	6.3
PBMC PHA-L	11.6	Lung fibroblast IL-4	27.2
Ramos (B cell) none	30.6	Lung fibroblast IL-9	26.8
Ramos (B cell)	**100.0**	Lung fibroblast IL-13	21.8

536

ionomycin			
B lymphocytes PWM	77.4	Lung fibroblast IFN gamma	29.5
B lymphocytes CD40L and IL-4	12.2	Dermal fibroblast CCD1070 rest	42.3
EOL-1 dbcAMP	13.0	Dermal fibroblast CCD1070 TNF alpha	51.4
EOL-1 dbcAMP PMA/ionomycin	6.9	Dermal fibroblast CCD1070 IL-1 beta	22.5
Dendritic cells none	4.5	Dermal fibroblast IFN gamma	11.1
Dendritic cells LPS	3.8	Dermal fibroblast IL-4	19.5
Dendritic cells anti-CD40	2.9	IBD Colitis 2	0.7
Monocytes rest	2.2	IBD Crohn's	0.9
Monocytes LPS	1.3	Colon	7.6
Macrophages rest	6.6	Lung	6.2
Macrophages LPS	2.7	Thymus	9.4
HUVEC none	17.4	Kidney	4.2
HUVEC starved	37.4		

CNS_neurodegeneration_v1.0 Summary: Ag3375 This panel does not show differential expression of the CG58572-01 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

5 **Panel 1.3D Summary:** Ag3375 - This gene is expressed at moderate to low levels in all samples on this panel, with the highest expression in gastric cancer cell line NCI-N87 (CT=28.8). Based on expression in this panel, this gene may be involved in gastric, pancreatic, brain, colon, renal, lung, breast, ovarian and prostate cancer as well as melanomas. Thus, expression of this gene could be used as a diagnostic marker for the

10 presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene might be of use in the treatment of these cancers.

This gene product is also expressed in adipose, pancreas, adrenal, thyroid, pituitary, skeletal muscle, heart, and liver. This widespread expression in tissues with metabolic function suggests that this gene product may be important for the pathogenesis, diagnosis,

15 and/or treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

537

In addition, this gene is expressed at moderate levels in the CNS. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4D Summary: Ag3375 The CG58572-01 gene is ubiquitously expressed on this panel, with highest expression in the B cell line Ramos treated with ionomycin (CT=26.2). Significant levels of expression are also seen in pokeweed mitogen-activated B lymphocytes. Therefore, therapies that antagonize the function of this gene product may be useful as therapeutic drugs to reduce or eliminate the symptoms in patients with autoimmune and inflammatory diseases in which B cells play a part in the initiation or progression of the disease process, such as lupus erythematosus, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, or psoriasis.

Interestingly, there is a difference between the levels of expression in resting and activated secondary T cells. The level in activated secondary T cells (CT=28.7-29.2) appears to be higher than in resting T cells (CT=31.3-33.1). Therefore, therapeutics designed with the protein encoded by this transcript could be important in the regulation of T cell function.

L. CG58564-01 and CG58564-02: PROTEIN TYROSINE PHOSPHATASE -

Expression of gene CG58564-01 and full length clone CG58564-02 was assessed using the primer-probe sets Ag3023 and Ag3373, described in Tables LA and LB. Results of the RTQ-PCR runs are shown in Tables LC, LD, LE and LF.

Table LA. Probe Name Ag3023

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctaatgctggatttgtccatca-3'	22	492	402
Probe	TET-5'-tcaggaatatgaagccatctacctagca-3'-TAMRA	28	517	403
Reverse	5'-tggagtggtgacatcatctgta-3'	22	555	404

Table LB. Probe Name Ag3373

Primers	Sequences	Length	Start Position	SEQ ID NO:

Forward	5'-atttgtccatcaacttcaggaa-3'	22	502	405
Probe	TET-5'-tgaagccatctacctagcaaaattaaca-3'-TAMRA	28	526	406
Reverse	5'-tggagtggtgacatcatctgta-3'	22	555	407

Table LC. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3023, Run 209821074	Rel. Exp.(%) Ag3373, Run 210154071	Tissue Name	Rel. Exp.(%) Ag3023, Run 209821074	Rel. Exp.(%) Ag3373, Run 210154071
AD 1 Hippo	10.9	16.8	Control (Path) 3 Temporal Ctx	9.1	8.0
AD 2 Hippo	34.2	37.6	Control (Path) 4 Temporal Ctx	40.6	65.5
AD 3 Hippo	12.0	15.8	AD 1 Occipital Ctx	24.7	29.1
AD 4 Hippo	13.8	10.3	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	60.7	57.8	AD 3 Occipital Ctx	14.7	15.0
AD 6 Hippo	80.7	72.2	AD 4 Occipital Ctx	35.4	22.4
Control 2 Hippo	35.8	38.4	AD 5 Occipital Ctx	3.9	30.4
Control 4 Hippo	16.5	11.7	AD 6 Occipital Ctx	46.0	37.4
Control (Path) 3 Hippo	13.1	15.4	Control 1 Occipital Ctx	9.9	10.7
AD 1 Temporal Ctx	39.0	31.4	Control 2 Occipital Ctx	39.0	38.4
AD 2 Temporal Ctx	38.7	73.2	Control 3 Occipital Ctx	23.0	20.6
AD 3 Temporal	9.5	13.2	Control 4	13.3	13.3

539

Ctx			Occipital Ctx		
AD 4 Temporal Ctx	27.9	34.9	Control (Path) 1 Occipital Ctx	80.1	76.3
AD 5 Inf Temporal Ctx	59.0	**100.0**	Control (Path) 2 Occipital Ctx	17.3	20.0
AD 5 SupTemporal Ctx	33.2	44.1	Control (Path) 3 Occipital Ctx	8.4	8.7
AD 6 Inf Temporal Ctx	**100.0**	73.2	Control (Path) 4 Occipital Ctx	21.2	20.6
AD 6 Sup Temporal Ctx	79.6	80.1	Control 1 Parietal Ctx	12.1	16.3
Control 1 Temporal Ctx	10.2	13.7	Control 2 Parietal Ctx	48.0	40.9
Control 2 Temporal Ctx	41.2	31.9	Control 3 Parietal Ctx	17.9	16.3
Control 3 Temporal Ctx	20.3	20.0	Control (Path) 1 Parietal Ctx	74.7	64.2
Control 4 Temporal Ctx	9.7	9.9	Control (Path) 2 Parietal Ctx	28.9	59.9
Control (Path) 1 Temporal Ctx	59.9	68.3	Control (Path) 3 Parietal Ctx	10.2	9.0
Control (Path) 2 Temporal Ctx	40.3	41.2	Control (Path) 4 Parietal Ctx	44.8	43.8

Table LD. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3373, Run 217043119	Tissue Name	Rel. Exp.(%) Ag3373, Run 217043119
Adipose	12.0	Renal ca. TK-10	20.3
Melanoma* Hs688(A).T	30.8	Bladder	23.2
Melanoma* Hs688(B).T	69.3	Gastric ca. (liver met.) NCI-N87	25.3

Melanoma* M14	15.0	Gastric ca. KATO III	30.8
Melanoma* LOXIMVI	26.6	Colon ca. SW-948	9.7
Melanoma* SK-MEL-5	21.5	Colon ca. SW480	35.1
Squamous cell carcinoma SCC-4	33.0	Colon ca.* (SW480 met) SW620	13.9
Testis Pool	19.8	Colon ca. HT29	8.5
Prostate ca.* (bone met) PC-3	**100.0**	Colon ca. HCT-116	36.9
Prostate Pool	9.2	Colon ca. CaCo-2	42.9
Placenta	3.8	Colon cancer tissue	9.0
Uterus Pool	7.4	Colon ca. SW1116	5.8
Ovarian ca. OVCAR-3	28.5	Colon ca. Colo-205	4.3
Ovarian ca. SK-OV-3	40.3	Colon ca. SW-48	4.2
Ovarian ca. OVCAR-4	20.0	Colon Pool	20.7
Ovarian ca. OVCAR-5	35.1	Small Intestine Pool	12.2
Ovarian ca. IGROV-1	10.9	Stomach Pool	9.9
Ovarian ca. OVCAR-8	9.2	Bone Marrow Pool	11.6
Ovary	9.7	Fetal Heart	20.7
Breast ca. MCF-7	37.6	Heart Pool	10.6
Breast ca. MDA-MB-231	37.1	Lymph Node Pool	17.9
Breast ca. BT 549	62.4	Fetal Skeletal Muscle	12.3
Breast ca. T47D	61.1	Skeletal Muscle Pool	16.0
Breast ca. MDA-N	10.0	Spleen Pool	11.6
Breast Pool	17.3	Thymus Pool	12.2
Trachea	12.0	CNS cancer (glio/astro) U87-MG	29.1
Lung	6.7	CNS cancer (glio/astro) U-118-MG	69.3
Fetal Lung	34.2	CNS cancer (neuro;met) SK-N-AS	34.9
Lung ca. NCI-N417	5.4	CNS cancer (astro) SF-539	19.1
Lung ca. LX-1	17.2	CNS cancer (astro) SNB-75	35.8
Lung ca. NCI-H146	3.0	CNS cancer (glio) SNB-19	11.3

Lung ca. SHP-77	18.6	CNS cancer (glio) SF-295	26.4
Lung ca. A549	29.1	Brain (Amygdala) Pool	4.5
Lung ca. NCI-H526	4.6	Brain (cerebellum)	8.1
Lung ca. NCI-H23	31.6	Brain (fetal)	13.2
Lung ca. NCI-H460	18.2	Brain (Hippocampus) Pool	5.3
Lung ca. HOP-62	14.1	Cerebral Cortex Pool	5.4
Lung ca. NCI-H522	31.6	Brain (Substantia nigra) Pool	4.8
Liver	1.2	Brain (Thalamus) Pool	8.0
Fetal Liver	32.3	Brain (whole)	6.2
Liver ca. HepG2	14.6	Spinal Cord Pool	6.6
Kidney Pool	22.1	Adrenal Gland	8.1
Fetal Kidney	26.1	Pituitary gland Pool	3.0
Renal ca. 786-0	28.7	Salivary Gland	4.7
Renal ca. A498	11.3	Thyroid (female)	4.4
Renal ca. ACHN	12.2	Pancreatic ca. CAPAN2	17.3
Renal ca. UO-31	24.1	Pancreas Pool	17.1

Table LE. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3023, Run 167966931	Tissue Name	Rel. Exp.(%) Ag3023, Run 167966931
Liver adenocarcinoma	51.1	Kidney (fetal)	26.2
Pancreas	6.1	Renal ca. 786-0	34.2
Pancreatic ca. CAPAN 2	17.7	Renal ca. A498	17.6
Adrenal gland	3.8	Renal ca. RXF 393	17.2
Thyroid	3.0	Renal ca. ACHN	13.5
Salivary gland	3.9	Renal ca. UO-31	0.0
Pituitary gland	3.6	Renal ca. TK-10	23.0
Brain (fetal)	8.1	Liver	11.7
Brain (whole)	8.5	Liver (fetal)	8.0
Brain (amygdala)	6.7	Liver ca. (hepatoblast) HepG2	26.2
Brain (cerebellum)	15.2	Lung	3.1
Brain (hippocampus)	5.4	Lung (fetal)	11.0
Brain (substantia nigra)	9.0	Lung ca. (small cell) LX-1	12.9
Brain (thalamus)	4.2	Lung ca. (small cell)	9.9

542

		NCI-H69	
Cerebral Cortex	2.0	Lung ca. (s.cell var.) SHP-77	67.8
Spinal cord	6.9	Lung ca. (large cell)NCI-H460	3.4
glio/astro U87-MG	28.5	Lung ca. (non-sm. cell) A549	45.1
glio/astro U-118-MG	46.7	Lung ca. (non-s.cell) NCI-H23	22.7
astrocytoma SW1783	40.6	Lung ca. (non-s.cell) HOP-62	25.7
neuro*; met SK-N-AS	27.2	Lung ca. (non-s.cl) NCI-H522	38.2
astrocytoma SF-539	29.7	Lung ca. (squam.) SW 900	27.4
astrocytoma SNB-75	35.1	Lung ca. (squam.) NCI-H596	29.9
glioma SNB-19	15.6	Mammary gland	5.1
glioma U251	37.9	Breast ca.* (pl.ef) MCF-7	47.0
glioma SF-295	18.4	Breast ca.* (pl.ef) MDA-MB-231	22.7
Heart (fetal)	2.9	Breast ca.* (pl.ef) T47D	86.5
Heart	12.9	Breast ca. BT-549	15.9
Skeletal muscle (fetal)	3.4	Breast ca. MDA-N	10.4
Skeletal muscle	36.3	Ovary	2.9
Bone marrow	4.5	Ovarian ca. OVCAR-3	26.1
Thymus	14.3	Ovarian ca. OVCAR-4	16.3
Spleen	8.7	Ovarian ca. OVCAR-5	83.5
Lymph node	11.8	Ovarian ca. OVCAR-8	9.3
Colorectal	10.4	Ovarian ca. IGROV-1	12.0
Stomach	7.8	Ovarian ca.* (ascites) SK-OV-3	**100.0**
Small intestine	5.1	Uterus	4.9
Colon ca. SW480	19.3	Placenta	1.3
Colon ca.* SW620(SW480 met)	42.9	Prostate	3.9
Colon ca. HT29	9.9	Prostate ca.* (bone met)PC-3	78.5

Colon ca. HCT-116	26.2	Testis	9.7
Colon ca. CaCo-2	41.5	Melanoma Hs688(A).T	5.9
Colon ca. tissue(ODO3866)	6.3	Melanoma* (met) Hs688(B).T	14.2
Colon ca. HCC-2998	16.0	Melanoma UACC-62	14.0
Gastric ca.* (liver met) NCI-N87	18.8	Melanoma M14	5.7
Bladder	30.6	Melanoma LOX IMVI	8.8
Trachea	3.2	Melanoma* (met) SK-MEL-5	14.7
Kidney	9.6	Adipose	18.9

<u>Table LF</u>. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3023, Run 164516146	Rel. Exp.(%) Ag3373, Run 165296617	Tissue Name	Rel. Exp.(%) Ag3023, Run 164516146	Rel. Exp.(%) Ag3373, Run 165296617
Secondary Th1 act	18.6	17.9	HUVEC IL-1beta	20.3	18.6
Secondary Th2 act	24.3	28.5	HUVEC IFN gamma	25.3	22.7
Secondary Tr1 act	22.8	21.8	HUVEC TNF alpha + IFN gamma	16.3	18.0
Secondary Th1 rest	7.5	6.8	HUVEC TNF alpha + IL4	18.2	13.4
Secondary Th2 rest	11.6	9.5	HUVEC IL-11	13.7	9.9
Secondary Tr1 rest	12.1	10.7	Lung Microvascular EC none	25.7	21.6
Primary Th1 act	20.7	16.5	Lung Microvascular EC TNFalpha + IL-1beta	26.2	18.3
Primary Th2 act	20.2	19.3	Microvascular Dermal EC none	27.5	21.3
Primary Tr1 act	23.3	27.7	Microsvasular Dermal EC TNFalpha + IL-1beta	20.7	19.9
Primary Th1 rest	51.1	51.4	Bronchial epithelium	13.0	16.3

544

			TNFalpha + IL1beta		
Primary Th2 rest	26.2	29.5	Small airway epithelium none	8.1	8.5
Primary Tr1 rest	23.7	26.1	Small airway epithelium TNFalpha + IL-1beta	50.3	39.8
CD45RA CD4 lymphocyte act	14.6	11.0	Coronery artery SMC rest	20.2	18.9
CD45RO CD4 lymphocyte act	25.2	22.4	Coronery artery SMC TNFalpha + IL-1beta	12.0	9.8
CD8 lymphocyte act	20.4	15.8	Astrocytes rest	10.4	11.1
Secondary CD8 lymphocyte rest	16.5	19.9	Astrocytes TNFalpha + IL-1beta	11.7	9.8
Secondary CD8 lymphocyte act	13.2	9.3	KU-812 (Basophil) rest	47.6	38.2
CD4 lymphocyte none	17.1	11.6	KU-812 (Basophil) PMA/ionomycin	94.0	92.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	18.3	16.6	CCD1106 (Keratinocytes) none	19.9	13.2
LAK cells rest	25.5	16.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.0	4.8
LAK cells IL-2	27.2	22.5	Liver cirrhosis	3.1	2.7
LAK cells IL-2+IL-12	27.2	19.3	Lupus kidney	2.1	1.7
LAK cells IL-2+IFN gamma	36.3	34.4	NCI-H292 none	30.1	18.9
LAK cells IL-2+IL-18	35.1	29.7	NCI-H292 IL-4	33.9	34.6
LAK cells PMA/ionomycin	12.4	11.0	NCI-H292 IL-9	40.1	29.1
NK Cells IL-2 rest	20.0	15.0	NCI-H292 IL-13	16.2	14.2
Two Way MLR 3 day	24.0	16.7	NCI-H292 IFN gamma	16.6	18.4
Two Way MLR 5 day	12.9	10.1	HPAEC none	13.6	13.5
Two Way MLR 7 day	11.4	9.5	HPAEC TNF alpha + IL-1 beta	25.3	25.3

PBMC rest	13.7	10.5	Lung fibroblast none	11.4	14.2
PBMC PWM	69.3	66.4	Lung fibroblast TNF alpha + IL-1 beta	6.1	7.2
PBMC PHA-L	22.8	17.7	Lung fibroblast IL-4	28.5	29.1
Ramos (B cell) none	24.1	19.3	Lung fibroblast IL-9	23.0	23.3
Ramos (B cell) ionomycin	**100.0**	**100.0**	Lung fibroblast IL-13	20.6	18.9
B lymphocytes PWM	71.7	74.2	Lung fibroblast IFN gamma	39.0	32.5
B lymphocytes CD40L and IL-4	29.1	28.7	Dermal fibroblast CCD1070 rest	33.9	31.0
EOL-1 dbcAMP	12.1	10.5	Dermal fibroblast CCD1070 TNF alpha	76.8	62.0
EOL-1 dbcAMP PMA/ionomycin	14.5	10.9	Dermal fibroblast CCD1070 IL-1 beta	20.3	13.9
Dendritic cells none	13.2	14.8	Dermal fibroblast IFN gamma	14.2	9.5
Dendritic cells LPS	11.7	8.3	Dermal fibroblast IL-4	26.4	20.4
Dendritic cells anti-CD40	17.7	12.7	IBD Colitis 2	2.6	2.2
Monocytes rest	16.7	17.6	IBD Crohn's	2.0	1.9
Monocytes LPS	6.4	5.0	Colon	11.9	10.5
Macrophages rest	23.5	22.8	Lung	13.3	11.2
Macrophages LPS	9.9	7.1	Thymus	14.4	12.9
HUVEC none	20.6	17.9	Kidney	27.5	19.6
HUVEC starved	43.5	38.4			

CNS_neurodegeneration_v1.0 Summary: Ag3023/Ag3373 This panel does not show differential expression of the CG58564-01 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

5 **General_screening_panel_v1.4 Summary:** Ag3373 Highest expression of the CG58564-01 gene is seen in a prostate cancer cell line (CT=27). Overall, this gene is expressed at moderate levels in the cancer cell lines in this panel. A higher level of expression is observed in clusters of cell lines derived from prostate, brain, melanoma, colon, lung,

546

breast and ovarian cancer when compared to expression in normal prostate, brain, colon, lung, breast and ovary. Thus, this gene could potentially be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be effective in the treatment of cancer in these tissues.

5   Among tissues with metabolic function, this gene product has moderate levels of expression in adipose, heart, skeletal muscle, adrenal, pituitary, thyroid and pancreas. Thus, this gene product may be a small molecule target for the treatment of endocrine and metabolic diseases, including obesity and Types 1 and 2 diabetes.

In addition, this gene appears to be differentially expressed in fetal (CT value = 29)
10   vs adult liver (CT value =33) and may be useful for differentiation between the two sources of this tissue.

This gene is also expressed at moderate levels in all central nervous system samples present on this panel. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

15   **Panel 1.3D Summary:** Ag3023 The CG58564-01 gene is ubiquitously expressed among the samples on this panel, with highest expression in an ovarian cancer cell line (CT=28.8). Overall, the expression of this gene shows good agreement with panel 1.4. A higher level of expression is observed in prostate, brain, melanoma, colon, lung, pancreatic, breast and ovarian cancer cell lines than the normal prostate, brain, colon, lung, pancreas, breast and
20   ovary. Thus, expression of this gene could be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be effective in the treatment of cancer in these tissues.

Among tissues with metabolic function, expression of this gene is widespread, as in the previous panel. Please see Panel 1.4 for discussion of utility of this gene in metabolic
25   disease.

This gene represents a phosphatase that is also expressed at low to moderate levels across the CNS. Some phosphatases comprise a family of MAP kinase regulating enzymes, members of which are upregulated in brains subjected to insults such as ischemia and seizure activity. MAP kinases are kown to regulate neurotrophic and neurotoxic pathways.

547

Consequently, agents that modulate the activity of this gene may have utility in attenuating the apoptotic and neurodegenerative processes following brain insults.

References:

1. Wiessner C. The dual specificity phosphatase PAC-1 is transcriptionally induced in the rat brain following transient forebrain ischemia. Brain Res Mol Brain Res 1995 Feb;28(2):353-6

2. Boschert U, Muda M, Camps M, Dickinson R, Arkinstall S. Induction of the dual specificity phosphatase PAC1 in rat brain following seizure activity. Neuroreport 1997 Sep 29;8(14):3077-80

Panel 4D Summary: Ag3023/Ag3373 The CG585864-01 gene is expressed at high to moderate levels in a wide range of cell types and tissues of significance in the immune response in health and disease. Highest expression of this gene is seen in ionomycin treated Ramos B cells (CT=26.83). Therefore, targeting of this gene product with a small molecule drug or antibody therapeutic may modulate the functions of cells of the immune system as well as resident tissue cells and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, and arthritis, including osteoarthritis and rheumatoid arthritis.

M. CG58564-03: Dual specificity phosphatase

Expression of gene CG58564-03 was assessed using the primer-probe sets Ag3023, Ag3373 and Ag5847, described in Tables MA, MB and MC. Results of the RTQ-PCR runs are shown in Tables MD, ME, MF, MG and MH.

Table MA. Probe Name Ag3023

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctaatgctggatttgtccatca-3'	22	261	408
Probe	TET-5'-tcaggaatatgaagccatctacctagca-3'-TAMRA	28	230	409
Reverse	5'-tggagtggtgacatcatctgta-3'	22	198	410

Table MB. Probe Name Ag3373

548

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atttgtccatcaacttcaggaa-3'	22	251	411
Probe	TET-5'-tgaagccatctacctagcaaaattaaca-3'-TAMRA	28	221	412
Reverse	5'-tggagtggtgacatcatctgta-3'	22	198	413

Table MC. Probe Name Ag5847

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cattccaaatgtttctgtagt-3'	21	335	414
Probe	TET-5'-ttcatagcagatgaatatgggcctaagaac-3'-TAMRA	30	371	415
Reverse	5'-ccacagtgcaaggaagac-3'	18	457	416

Table MD. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3023, Run 209821074	Rel. Exp.(%) Ag3373, Run 210154071	Tissue Name	Rel. Exp.(%) Ag3023, Run 209821074	Rel. Exp.(%) Ag3373, Run 210154071
AD 1 Hippo	10.9	16.8	Control (Path) 3 Temporal Ctx	9.1	8.0
AD 2 Hippo	34.2	37.6	Control (Path) 4 Temporal Ctx	40.6	65.5
AD 3 Hippo	12.0	15.8	AD 1 Occipital Ctx	24.7	29.1
AD 4 Hippo	13.8	10.3	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	60.7	57.8	AD 3 Occipital Ctx	14.7	15.0
AD 6 Hippo	80.7	72.2	AD 4 Occipital Ctx	35.4	22.4
Control 2 Hippo	35.8	38.4	AD 5 Occipital Ctx	3.9	30.4
Control 4	16.5	11.7	AD 6	46.0	37.4

Hippo			Occipital Ctx		
Control (Path) 3 Hippo	13.1	15.4	Control 1 Occipital Ctx	9.9	10.7
AD 1 Temporal Ctx	39.0	31.4	Control 2 Occipital Ctx	39.0	38.4
AD 2 Temporal Ctx	38.7	73.2	Control 3 Occipital Ctx	23.0	20.6
AD 3 Temporal Ctx	9.5	13.2	Control 4 Occipital Ctx	13.3	13.3
AD 4 Temporal Ctx	27.9	34.9	Control (Path) 1 Occipital Ctx	80.1	76.3
AD 5 Inf Temporal Ctx	59.0	**100.0**	Control (Path) 2 Occipital Ctx	17.3	20.0
AD 5 SupTemporal Ctx	33.2	44.1	Control (Path) 3 Occipital Ctx	8.4	8.7
AD 6 Inf Temporal Ctx	**100.0**	73.2	Control (Path) 4 Occipital Ctx	21.2	20.6
AD 6 Sup Temporal Ctx	79.6	80.1	Control 1 Parietal Ctx	12.1	16.3
Control 1 Temporal Ctx	10.2	13.7	Control 2 Parietal Ctx	48.0	40.9
Control 2 Temporal Ctx	41.2	31.9	Control 3 Parietal Ctx	17.9	16.3
Control 3 Temporal Ctx	20.3	20.0	Control (Path) 1 Parietal Ctx	74.7	64.2
Control 4 Temporal Ctx	9.7	9.9	Control (Path) 2 Parietal Ctx	28.9	59.9
Control (Path) 1 Temporal Ctx	59.9	68.3	Control (Path) 3 Parietal Ctx	10.2	9.0
Control (Path) 2 Temporal Ctx	40.3	41.2	Control (Path) 4	44.8	43.8

			Parietal Ctx		

Table ME. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3373, Run 217043119	Tissue Name	Rel. Exp.(%) Ag3373, Run 217043119
Adipose	12.0	Renal ca. TK-10	20.3
Melanoma* Hs688(A).T	30.8	Bladder	23.2
Melanoma* Hs688(B).T	69.3	Gastric ca. (liver met.) NCI-N87	25.3
Melanoma* M14	15.0	Gastric ca. KATO III	30.8
Melanoma* LOXIMVI	26.6	Colon ca. SW-948	9.7
Melanoma* SK-MEL-5	21.5	Colon ca. SW480	35.1
Squamous cell carcinoma SCC-4	33.0	Colon ca.* (SW480 met) SW620	13.9
Testis Pool	19.8	Colon ca. HT29	8.5
Prostate ca.* (bone met) PC-3	100.0	Colon ca. HCT-116	36.9
Prostate Pool	9.2	Colon ca. CaCo-2	42.9
Placenta	3.8	Colon cancer tissue	9.0
Uterus Pool	7.4	Colon ca. SW1116	5.8
Ovarian ca. OVCAR-3	28.5	Colon ca. Colo-205	4.3
Ovarian ca. SK-OV-3	40.3	Colon ca. SW-48	4.2
Ovarian ca. OVCAR-4	20.0	Colon Pool	20.7
Ovarian ca. OVCAR-5	35.1	Small Intestine Pool	12.2
Ovarian ca. IGROV-1	10.9	Stomach Pool	9.9
Ovarian ca. OVCAR-8	9.2	Bone Marrow Pool	11.6
Ovary	9.7	Fetal Heart	20.7
Breast ca. MCF-7	37.6	Heart Pool	10.6
Breast ca. MDA-MB-231	37.1	Lymph Node Pool	17.9
Breast ca. BT 549	62.4	Fetal Skeletal Muscle	12.3
Breast ca. T47D	61.1	Skeletal Muscle Pool	16.0
Breast ca. MDA-N	10.0	Spleen Pool	11.6

Breast Pool	17.3	Thymus Pool	12.2
Trachea	12.0	CNS cancer (glio/astro) U87-MG	29.1
Lung	6.7	CNS cancer (glio/astro) U-118-MG	69.3
Fetal Lung	34.2	CNS cancer (neuro;met) SK-N-AS	34.9
Lung ca. NCI-N417	5.4	CNS cancer (astro) SF-539	19.1
Lung ca. LX-1	17.2	CNS cancer (astro) SNB-75	35.8
Lung ca. NCI-H146	3.0	CNS cancer (glio) SNB-19	11.3
Lung ca. SHP-77	18.6	CNS cancer (glio) SF-295	26.4
Lung ca. A549	29.1	Brain (Amygdala) Pool	4.5
Lung ca. NCI-H526	4.6	Brain (cerebellum)	8.1
Lung ca. NCI-H23	31.6	Brain (fetal)	13.2
Lung ca. NCI-H460	18.2	Brain (Hippocampus) Pool	5.3
Lung ca. HOP-62	14.1	Cerebral Cortex Pool	5.4
Lung ca. NCI-H522	31.6	Brain (Substantia nigra) Pool	4.8
Liver	1.2	Brain (Thalamus) Pool	8.0
Fetal Liver	32.3	Brain (whole)	6.2
Liver ca. HepG2	14.6	Spinal Cord Pool	6.6
Kidney Pool	22.1	Adrenal Gland	8.1
Fetal Kidney	26.1	Pituitary gland Pool	3.0
Renal ca. 786-0	28.7	Salivary Gland	4.7
Renal ca. A498	11.3	Thyroid (female)	4.4
Renal ca. ACHN	12.2	Pancreatic ca. CAPAN2	17.3
Renal ca. UO-31	24.1	Pancreas Pool	17.1

Table MF. General_screening_panel_v1.5

Tissue Name	Rel. Exp.(%) Ag5847, Run 247590257	Tissue Name	Rel. Exp.(%) Ag5847, Run 247590257
Adipose	0.1	Renal ca. TK-10	0.2
Melanoma* Hs688(A).T	0.1	Bladder	0.1
Melanoma* Hs688(B).T	0.1	Gastric ca. (liver met.) NCI-N87	0.2

Melanoma* M14	0.1	Gastric ca. KATO III	0.1
Melanoma* LOXIMVI	0.1	Colon ca. SW-948	0.1
Melanoma* SK-MEL-5	0.1	Colon ca. SW480	0.2
Squamous cell carcinoma SCC-4	0.2	Colon ca.* (SW480 met) SW620	1.8
Testis Pool	0.1	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.6	Colon ca. HCT-116	0.2
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.2	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.1	Colon Pool	0.1
Ovarian ca. OVCAR-5	0.2	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.1	Bone Marrow Pool	0.0
Ovary	0.1	Fetal Heart	0.1
Breast ca. MCF-7	0.3	Heart Pool	0.0
Breast ca. MDA-MB-231	0.2	Lymph Node Pool	0.1
Breast ca. BT 549	0.2	Fetal Skeletal Muscle	0.1
Breast ca. T47D	0.2	Skeletal Muscle Pool	0.1
Breast ca. MDA-N	0.1	Spleen Pool	0.1
Breast Pool	0.0	Thymus Pool	0.1
Trachea	0.1	CNS cancer (glio/astro) U87-MG	0.2
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.5
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	0.2
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.1
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.2
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.1

Lung ca. SHP-77	0.1	CNS cancer (glio) SF-295	0.2
Lung ca. A549	0.2	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.1	Brain (fetal)	0.1
Lung ca. NCI-H460	0.1	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.1	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.1	Brain (whole)	0.0
Liver ca. HepG2	0.1	Spinal Cord Pool	0.0
Kidney Pool	0.1	Adrenal Gland	0.0
Fetal Kidney	0.1	Pituitary gland Pool	0.0
Renal ca. 786-0	0.2	Salivary Gland	**100.0**
Renal ca. A498	0.1	Thyroid (female)	0.0
Renal ca. ACHN	0.1	Pancreatic ca. CAPAN2	0.1
Renal ca. UO-31	0.1	Pancreas Pool	0.0

Table MG. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3023, Run 167966931	Tissue Name	Rel. Exp.(%) Ag3023, Run 167966931
Liver adenocarcinoma	51.1	Kidney (fetal)	26.2
Pancreas	6.1	Renal ca. 786-0	34.2
Pancreatic ca. CAPAN 2	17.7	Renal ca. A498	17.6
Adrenal gland	3.8	Renal ca. RXF 393	17.2
Thyroid	3.0	Renal ca. ACHN	13.5
Salivary gland	3.9	Renal ca. UO-31	0.0
Pituitary gland	3.6	Renal ca. TK-10	23.0
Brain (fetal)	8.1	Liver	11.7
Brain (whole)	8.5	Liver (fetal)	8.0
Brain (amygdala)	6.7	Liver ca. (hepatoblast) HepG2	26.2
Brain (cerebellum)	15.2	Lung	3.1
Brain (hippocampus)	5.4	Lung (fetal)	11.0
Brain (substantia nigra)	9.0	Lung ca. (small cell) LX-1	12.9
Brain (thalamus)	4.2	Lung ca. (small cell)	9.9

554

		NCI-H69	
Cerebral Cortex	2.0	Lung ca. (s.cell var.) SHP-77	67.8
Spinal cord	6.9	Lung ca. (large cell)NCI-H460	3.4
Glio/astro U87-MG	28.5	Lung ca. (non-sm. cell) A549	45.1
Glio/astro U-118-MG	46.7	Lung ca. (non-s.cell) NCI-H23	22.7
astrocytoma SW1783	40.6	Lung ca. (non-s.cell) HOP-62	25.7
neuro*; met SK-N-AS	27.2	Lung ca. (non-s.cl) NCI-H522	38.2
astrocytoma SF-539	29.7	Lung ca. (squam.) SW 900	27.4
astrocytoma SNB-75	35.1	Lung ca. (squam.) NCI-H596	29.9
glioma SNB-19	15.6	Mammary gland	5.1
glioma U251	37.9	Breast ca.* (pl.ef) MCF-7	47.0
glioma SF-295	18.4	Breast ca.* (pl.ef) MDA-MB-231	22.7
Heart (fetal)	2.9	Breast ca.* (pl.ef) T47D	86.5
Heart	12.9	Breast ca. BT-549	15.9
Skeletal muscle (fetal)	3.4	Breast ca. MDA-N	10.4
Skeletal muscle	36.3	Ovary	2.9
Bone marrow	4.5	Ovarian ca. OVCAR-3	26.1
Thymus	14.3	Ovarian ca. OVCAR-4	16.3
Spleen	8.7	Ovarian ca. OVCAR-5	83.5
Lymph node	11.8	Ovarian ca. OVCAR-8	9.3
Colorectal	10.4	Ovarian ca. IGROV-1	12.0
Stomach	7.8	Ovarian ca.* (ascites) SK-OV-3	**100.0**
Small intestine	5.1	Uterus	4.9
Colon ca. SW480	19.3	Placenta	1.3
Colon ca.* SW620(SW480 met)	42.9	Prostate	3.9
Colon ca. HT29	9.9	Prostate ca.* (bone met)PC-3	78.5

555

Colon ca. HCT-116	26.2	Testis	9.7
Colon ca. CaCo-2	41.5	Melanoma Hs688(A).T	5.9
Colon ca. tissue(ODO3866)	6.3	Melanoma* (met) Hs688(B).T	14.2
Colon ca. HCC-2998	16.0	Melanoma UACC-62	14.0
Gastric ca.* (liver met) NCI-N87	18.8	Melanoma M14	5.7
Bladder	30.6	Melanoma LOX IMVI	8.8
Trachea	3.2	Melanoma* (met) SK-MEL-5	14.7
Kidney	9.6	Adipose	18.9

<u>Table MH</u>. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3023, Run 164516146	Rel. Exp.(%) Ag3373, Run 165296617	Tissue Name	Rel. Exp.(%) Ag3023, Run 164516146	Rel. Exp.(%) Ag3373, Run 165296617
Secondary Th1 act	18.6	17.9	HUVEC IL-1beta	20.3	18.6
Secondary Th2 act	24.3	28.5	HUVEC IFN gamma	25.3	22.7
Secondary Tr1 act	22.8	21.8	HUVEC TNF alpha + IFN gamma	16.3	18.0
Secondary Th1 rest	7.5	6.8	HUVEC TNF alpha + IL4	18.2	13.4
Secondary Th2 rest	11.6	9.5	HUVEC IL-11	13.7	9.9
Secondary Tr1 rest	12.1	10.7	Lung Microvascular EC none	25.7	21.6
Primary Th1 act	20.7	16.5	Lung Microvascular EC TNFalpha + IL-1beta	26.2	18.3
Primary Th2 act	20.2	19.3	Microvascular Dermal EC none	27.5	21.3
Primary Tr1 act	23.3	27.7	Microsvasular Dermal EC TNFalpha + IL-1beta	20.7	19.9
Primary Th1 rest	51.1	51.4	Bronchial epithelium	13.0	16.3

			TNFalpha + IL1beta		
Primary Th2 rest	26.2	29.5	Small airway epithelium none	8.1	8.5
Primary Tr1 rest	23.7	26.1	Small airway epithelium TNFalpha + IL-1beta	50.3	39.8
CD45RA CD4 lymphocyte act	14.6	11.0	Coronery artery SMC rest	20.2	18.9
CD45RO CD4 lymphocyte act	25.2	22.4	Coronery artery SMC TNFalpha + IL-1beta	12.0	9.8
CD8 lymphocyte act	20.4	15.8	Astrocytes rest	10.4	11.1
Secondary CD8 lymphocyte rest	16.5	19.9	Astrocytes TNFalpha + IL-1beta	11.7	9.8
Secondary CD8 lymphocyte act	13.2	9.3	KU-812 (Basophil) rest	47.6	38.2
CD4 lymphocyte none	17.1	11.6	KU-812 (Basophil) PMA/ionomycin	94.0	92.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	18.3	16.6	CCD1106 (Keratinocytes) none	19.9	13.2
LAK cells rest	25.5	16.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.0	4.8
LAK cells IL-2	27.2	22.5	Liver cirrhosis	3.1	2.7
LAK cells IL-2+IL-12	27.2	19.3	Lupus kidney	2.1	1.7
LAK cells IL-2+IFN gamma	36.3	34.4	NCI-H292 none	30.1	18.9
LAK cells IL-2+IL-18	35.1	29.7	NCI-H292 IL-4	33.9	34.6
LAK cells PMA/ionomycin	12.4	11.0	NCI-H292 IL-9	40.1	29.1
NK Cells IL-2 rest	20.0	15.0	NCI-H292 IL-13	16.2	14.2
Two Way MLR 3 day	24.0	16.7	NCI-H292 IFN gamma	16.6	18.4
Two Way MLR 5 day	12.9	10.1	HPAEC none	13.6	13.5
Two Way MLR 7 day	11.4	9.5	HPAEC TNF alpha + IL-1 beta	25.3	25.3

557

PBMC rest	13.7	10.5	Lung fibroblast none	11.4	14.2
PBMC PWM	69.3	66.4	Lung fibroblast TNF alpha + IL-1 beta	6.1	7.2
PBMC PHA-L	22.8	17.7	Lung fibroblast IL-4	28.5	29.1
Ramos (B cell) none	24.1	19.3	Lung fibroblast IL-9	23.0	23.3
Ramos (B cell) ionomycin	**100.0**	**100.0**	Lung fibroblast IL-13	20.6	18.9
B lymphocytes PWM	71.7	74.2	Lung fibroblast IFN gamma	39.0	32.5
B lymphocytes CD40L and IL-4	29.1	28.7	Dermal fibroblast CCD1070 rest	33.9	31.0
EOL-1 dbcAMP	12.1	10.5	Dermal fibroblast CCD1070 TNF alpha	76.8	62.0
EOL-1 dbcAMP PMA/ionomycin	14.5	10.9	Dermal fibroblast CCD1070 IL-1 beta	20.3	13.9
Dendritic cells none	13.2	14.8	Dermal fibroblast IFN gamma	14.2	9.5
Dendritic cells LPS	11.7	8.3	Dermal fibroblast IL-4	26.4	20.4
Dendritic cells anti-CD40	17.7	12.7	IBD Colitis 2	2.6	2.2
Monocytes rest	16.7	17.6	IBD Crohn's	2.0	1.9
Monocytes LPS	6.4	5.0	Colon	11.9	10.5
Macrophages rest	23.5	22.8	Lung	13.3	11.2
Macrophages LPS	9.9	7.1	Thymus	14.4	12.9
HUVEC none	20.6	17.9	Kidney	27.5	19.6
HUVEC starved	43.5	38.4			

CNS_neurodegeneration_v1.0 Summary: Ag3023/Ag3373 This panel does not show differential expression of the CG56804-03 gene, a splice variant of CG56804-01, in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system. Ag5847 - This primer pair recognizes only the splice variant CG58564-03. Expression of this variant is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

558

General_screening_panel_v1.4 Summary: Ag3373 Highest expression of the CG56804-03 gene is seen in a prostate cancer cell line (CT=27). Overall, this gene is expressed at moderate levels in the cancer cell lines in this panel. A higher level of expression is observed in clusters of cell lines derived from prostate, brain, melanoma, colon, lung,

5 breast and ovarian cancer when compared to expression in normal prostate, brain, colon, lung, breast and ovary. Thus, this gene could potentially be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be effective in the treatment of cancer in these tissues.

Among tissues with metabolic function, this gene product has moderate levels of

10 expression in adipose, heart, skeletal muscle, adrenal, pituitary, thyroid and pancreas. Thus, this gene product may be a small molecule target for the treatment of endocrine and metabolic diseases, including obesity and Types 1 and 2 diabetes.

In addition, this gene appears to be differentially expressed in fetal (CT value = 29) vs adult liver (CT value =33) and may be useful for differentiation between the two sources

15 of this tissue.

This gene is also expressed at moderate levels in all central nervous system samples present on this panel. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

General_screening_panel_v1.5 Summary: Ag5847 - This primer pair, specific to this

20 splice variant, CG58564-03. Expression of this variant is highest in salivary gland (CT=28.6). Therefore, expression of this gene can be used to differentiate this sample from others on the panel.

Panel 1.3D Summary: Ag3023 The CG56804-03 gene is ubiquitously expressed among the samples on this panel, with highest expression in an ovarian cancer cell line (CT=28.8).

25 Overall, the expression of this gene shows good agreement with panel 1.4. A higher level of expression is observed in prostate, brain, melanoma, colon, lung, pancreatic, breast and ovarian cancer cell lines than the normal prostate, brain, colon, lung, pancreas, breast and ovary. Thus, expression of this gene could be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule

30 drugs may be effective in the treatment of cancer in these tissues.

Among tissues with metabolic function, expression of this gene is widespread, as in the previous panel. Please see Panel 1.4 for discussion of utility of this gene in metabolic disease.

This gene represents a dual specificity phosphatase that is also expressed at low to moderate levels across the CNS. Dual-specificity phosphatases comprise a family of MAP kinase regulating enzymes, members of which are upregulated in brains subjected to insults such as ischemia and seizure activity. MAP kinases are kown to regulate neurotrophic and neurotoxic pathways. Consequently, agents that modulate the activity of this gene may have utility in attenuating the apoptotic and neurodegenerative processes following brain insults.

Panel 4.1D Summary: Ag5847 - This primer pair recognizes a splice variant of CG58564-03. Expression of this variant is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag3023/Ag3373 The CG56804-03 gene is expressed at high to moderate levels in a wide range of cell types and tissues of significance in the immune response in health and disease. Highest expression of this gene is seen in ionomycin treated Ramos B cells (CT=26.83). Therefore, targeting of this gene product with a small molecule drug or antibody therapeutic may modulate the functions of cells of the immune system as well as resident tissue cells and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, and arthritis, including osteoarthritis and rheumatoid arthritis.

N. CG58564-04: Dual specificity phosphatase

Expression of gene CG58564-04, a splice variant of CG58564-01, was assessed using the primer-probe sets Ag3023, Ag3373 and Ag5844, described in Tables NA, NB and NC. Results of the RTQ-PCR runs are shown in Tables ND, NE, NF and NG.

Table NA. Probe Name Ag3023

560

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctaatgctggatttgtccatca-3'	22	190	417
Probe	TET-5'-tcaggaatatgaagccatctacctagca-3'-TAMRA	28	159	418
Reverse	5'-tggagtggtgacatcatctgta-3'	22	127	419

Table NB. Probe Name Ag3373

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atttgtccatcaacttcaggaa-3'	22	180	420
Probe	TET-5'-tgaagccatctacctagcaaaattaaca-3'-TAMRA	28	150	421
Reverse	5'-tggagtggtgacatcatctgta-3'	22	127	422

Table NC. Probe Name Ag5844

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ccttagtctaaataactgctg-3'	21	377	423
Probe	TET-5'-agtttgcttcaatattttgtcgtatgcata-3'-TAMRA	30	415	424
Reverse	5'-aggagtggacctaccctat-3'	19	552	425

Table ND. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3023, Run 209821074	Rel. Exp.(%) Ag3373, Run 210154071	Tissue Name	Rel. Exp.(%) Ag3023, Run 209821074	Rel. Exp.(%) Ag3373, Run 210154071
AD 1 Hippo	10.9	16.8	Control (Path) 3 Temporal Ctx	9.1	8.0
AD 2 Hippo	34.2	37.6	Control (Path) 4 Temporal Ctx	40.6	65.5
AD 3 Hippo	12.0	15.8	AD 1 Occipital Ctx	24.7	29.1
AD 4 Hippo	13.8	10.3	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	60.7	57.8	AD 3	14.7	15.0

			Occipital Ctx		
AD 6 Hippo	80.7	72.2	AD 4 Occipital Ctx	35.4	22.4
Control 2 Hippo	35.8	38.4	AD 5 Occipital Ctx	3.9	30.4
Control 4 Hippo	16.5	11.7	AD 6 Occipital Ctx	46.0	37.4
Control (Path) 3 Hippo	13.1	15.4	Control 1 Occipital Ctx	9.9	10.7
AD 1 Temporal Ctx	39.0	31.4	Control 2 Occipital Ctx	39.0	38.4
AD 2 Temporal Ctx	38.7	73.2	Control 3 Occipital Ctx	23.0	20.6
AD 3 Temporal Ctx	9.5	13.2	Control 4 Occipital Ctx	13.3	13.3
AD 4 Temporal Ctx	27.9	34.9	Control (Path) 1 Occipital Ctx	80.1	76.3
AD 5 Inf Temporal Ctx	59.0	**100.0**	Control (Path) 2 Occipital Ctx	17.3	20.0
AD 5 SupTemporal Ctx	33.2	44.1	Control (Path) 3 Occipital Ctx	8.4	8.7
AD 6 Inf Temporal Ctx	**100.0**	73.2	Control (Path) 4 Occipital Ctx	21.2	20.6
AD 6 Sup Temporal Ctx	79.6	80.1	Control 1 Parietal Ctx	12.1	16.3
Control 1 Temporal Ctx	10.2	13.7	Control 2 Parietal Ctx	48.0	40.9
Control 2 Temporal Ctx	41.2	31.9	Control 3 Parietal Ctx	17.9	16.3
Control 3 Temporal Ctx	20.3	20.0	Control (Path) 1	74.7	64.2

			Parietal Ctx		
Control 4 Temporal Ctx	9.7	9.9	Control (Path) 2 Parietal Ctx	28.9	59.9
Control (Path) 1 Temporal Ctx	59.9	68.3	Control (Path) 3 Parietal Ctx	10.2	9.0
Control (Path) 2 Temporal Ctx	40.3	41.2	Control (Path) 4 Parietal Ctx	44.8	43.8

Table NE. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3373, Run 217043119	Tissue Name	Rel. Exp.(%) Ag3373, Run 217043119
Adipose	12.0	Renal ca. TK-10	20.3
Melanoma* Hs688(A).T	30.8	Bladder	23.2
Melanoma* Hs688(B).T	69.3	Gastric ca. (liver met.) NCI-N87	25.3
Melanoma* M14	15.0	Gastric ca. KATO III	30.8
Melanoma* LOXIMVI	26.6	Colon ca. SW-948	9.7
Melanoma* SK-MEL-5	21.5	Colon ca. SW480	35.1
Squamous cell carcinoma SCC-4	33.0	Colon ca.* (SW480 met) SW620	13.9
Testis Pool	19.8	Colon ca. HT29	8.5
Prostate ca.* (bone met) PC-3	100.0	Colon ca. HCT-116	36.9
Prostate Pool	9.2	Colon ca. CaCo-2	42.9
Placenta	3.8	Colon cancer tissue	9.0
Uterus Pool	7.4	Colon ca. SW1116	5.8
Ovarian ca. OVCAR-3	28.5	Colon ca. Colo-205	4.3
Ovarian ca. SK-OV-3	40.3	Colon ca. SW-48	4.2
Ovarian ca. OVCAR-4	20.0	Colon Pool	20.7
Ovarian ca. OVCAR-5	35.1	Small Intestine Pool	12.2
Ovarian ca. IGROV-1	10.9	Stomach Pool	9.9
Ovarian ca.	9.2	Bone Marrow Pool	11.6

OVCAR-8			
Ovary	9.7	Fetal Heart	20.7
Breast ca. MCF-7	37.6	Heart Pool	10.6
Breast ca. MDA-MB-231	37.1	Lymph Node Pool	17.9
Breast ca. BT 549	62.4	Fetal Skeletal Muscle	12.3
Breast ca. T47D	61.1	Skeletal Muscle Pool	16.0
Breast ca. MDA-N	10.0	Spleen Pool	11.6
Breast Pool	17.3	Thymus Pool	12.2
Trachea	12.0	CNS cancer (glio/astro) U87-MG	29.1
Lung	6.7	CNS cancer (glio/astro) U-118-MG	69.3
Fetal Lung	34.2	CNS cancer (neuro;met) SK-N-AS	34.9
Lung ca. NCI-N417	5.4	CNS cancer (astro) SF-539	19.1
Lung ca. LX-1	17.2	CNS cancer (astro) SNB-75	35.8
Lung ca. NCI-H146	3.0	CNS cancer (glio) SNB-19	11.3
Lung ca. SHP-77	18.6	CNS cancer (glio) SF-295	26.4
Lung ca. A549	29.1	Brain (Amygdala) Pool	4.5
Lung ca. NCI-H526	4.6	Brain (cerebellum)	8.1
Lung cà. NCI-H23	31.6	Brain (fetal)	13.2
Lung ca. NCI-H460	18.2	Brain (Hippocampus) Pool	5.3
Lung ca. HOP-62	14.1	Cerebral Cortex Pool	5.4
Lung ca. NCI-H522	31.6	Brain (Substantia nigra) Pool	4.8
Liver	1.2	Brain (Thalamus) Pool	8.0
Fetal Liver	32.3	Brain (whole)	6.2
Liver ca. HepG2	14.6	Spinal Cord Pool	6.6
Kidney Pool	22.1	Adrenal Gland	8.1
Fetal Kidney	26.1	Pituitary gland Pool	3.0
Renal ca. 786-0	28.7	Salivary Gland	4.7
Renal ca. A498	11.3	Thyroid (female)	4.4
Renal ca. ACHN	12.2	Pancreatic ca. CAPAN2	17.3
Renal ca. UO-31	24.1	Pancreas Pool	17.1

Table NF. Panel 1.3D

564

Tissue Name	Rel. Exp.(%) Ag3023, Run 167966931	Tissue Name	Rel. Exp.(%) Ag3023, Run 167966931
Liver adenocarcinoma	51.1	Kidney (fetal)	26.2
Pancreas	6.1	Renal ca. 786-0	34.2
Pancreatic ca. CAPAN 2	17.7	Renal ca. A498	17.6
Adrenal gland	3.8	Renal ca. RXF 393	17.2
Thyroid	3.0	Renal ca. ACHN	13.5
Salivary gland	3.9	Renal ca. UO-31	0.0
Pituitary gland	3.6	Renal ca. TK-10	23.0
Brain (fetal)	8.1	Liver	11.7
Brain (whole)	8.5	Liver (fetal)	8.0
Brain (amygdala)	6.7	Liver ca. (hepatoblast) HepG2	26.2
Brain (cerebellum)	15.2	Lung	3.1
Brain (hippocampus)	5.4	Lung (fetal)	11.0
Brain (substantia nigra)	9.0	Lung ca. (small cell) LX-1	12.9
Brain (thalamus)	4.2	Lung ca. (small cell) NCI-H69	9.9
Cerebral Cortex	2.0	Lung ca. (s.cell var.) SHP-77	67.8
Spinal cord	6.9	Lung ca. (large cell)NCI-H460	3.4
Glio/astro U87-MG	28.5	Lung ca. (non-sm. cell) A549	45.1
Glio/astro U-118-MG	46.7	Lung ca. (non-s.cell) NCI-H23	22.7
astrocytoma SW1783	40.6	Lung ca. (non-s.cell) HOP-62	25.7
neuro*; met SK-N-AS	27.2	Lung ca. (non-s.cl) NCI-H522	38.2
astrocytoma SF-539	29.7	Lung ca. (squam.) SW 900	27.4
astrocytoma SNB-75	35.1	Lung ca. (squam.) NCI-H596	29.9
glioma SNB-19	15.6	Mammary gland	5.1
glioma U251	37.9	Breast ca.* (pl.ef) MCF-7	47.0
glioma SF-295	18.4	Breast ca.* (pl.ef) MDA-MB-231	22.7
Heart (fetal)	2.9	Breast ca.* (pl.ef) T47D	86.5
Heart	12.9	Breast ca. BT-549	15.9

Skeletal muscle (fetal)	3.4	Breast ca. MDA-N	10.4
Skeletal muscle	36.3	Ovary	2.9
Bone marrow	4.5	Ovarian ca. OVCAR-3	26.1
Thymus	14.3	Ovarian ca. OVCAR-4	16.3
Spleen	8.7	Ovarian ca. OVCAR-5	83.5
Lymph node	11.8	Ovarian ca. OVCAR-8	9.3
Colorectal	10.4	Ovarian ca. IGROV-1	12.0
Stomach	7.8	Ovarian ca.* (ascites) SK-OV-3	**100.0**
Small intestine	5.1	Uterus	4.9
Colon ca. SW480	19.3	Placenta	1.3
Colon ca.* SW620(SW480 met)	42.9	Prostate	3.9
Colon ca. HT29	9.9	Prostate ca.* (bone met)PC-3	78.5
Colon ca. HCT-116	26.2	Testis	9.7
Colon ca. CaCo-2	41.5	Melanoma Hs688(A).T	5.9
Colon ca. tissue(ODO3866)	6.3	Melanoma* (met) Hs688(B).T	14.2
Colon ca. HCC-2998	16.0	Melanoma UACC-62	14.0
Gastric ca.* (liver met) NCI-N87	18.8	Melanoma M14	5.7
Bladder	30.6	Melanoma LOX IMVI	8.8
Trachea	3.2	Melanoma* (met) SK-MEL-5	14.7
Kidney	9.6	Adipose	18.9

Table NG. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3023, Run 164516146	Rel. Exp.(%) Ag3373, Run 165296617	Tissue Name	Rel. Exp.(%) Ag3023, Run 164516146	Rel. Exp.(%) Ag3373, Run 165296617
Secondary Th1 act	18.6	17.9	HUVEC IL-1beta	20.3	18.6
Secondary Th2 act	24.3	28.5	HUVEC IFN gamma	25.3	22.7

Secondary Tr1 act	22.8	21.8	HUVEC TNF alpha + IFN gamma	16.3	18.0
Secondary Th1 rest	7.5	6.8	HUVEC TNF alpha + IL4	18.2	13.4
Secondary Th2 rest	11.6	9.5	HUVEC IL-11	13.7	9.9
Secondary Tr1 rest	12.1	10.7	Lung Microvascular EC none	25.7	21.6
Primary Th1 act	20.7	16.5	Lung Microvascular EC TNFalpha + IL-1beta	26.2	18.3
Primary Th2 act	20.2	19.3	Microvascular Dermal EC none	27.5	21.3
Primary Tr1 act	23.3	27.7	Microsvasular Dermal EC TNFalpha + IL-1beta	20.7	19.9
Primary Th1 rest	51.1	51.4	Bronchial epithelium TNFalpha + IL1beta	13.0	16.3
Primary Th2 rest	26.2	29.5	Small airway epithelium none	8.1	8.5
Primary Tr1 rest	23.7	26.1	Small airway epithelium TNFalpha + IL-1beta	50.3	39.8
CD45RA CD4 lymphocyte act	14.6	11.0	Coronery artery SMC rest	20.2	18.9
CD45RO CD4 lymphocyte act	25.2	22.4	Coronery artery SMC TNFalpha + IL-1beta	12.0	9.8
CD8 lymphocyte act	20.4	15.8	Astrocytes rest	10.4	11.1
Secondary CD8 lymphocyte rest	16.5	19.9	Astrocytes TNFalpha + IL-1beta	11.7	9.8
Secondary CD8 lymphocyte act	13.2	9.3	KU-812 (Basophil) rest	47.6	38.2
CD4 lymphocyte none	17.1	11.6	KU-812 (Basophil) PMA/ionomycin	94.0	92.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	18.3	16.6	CCD1106 (Keratinocytes) none	19.9	13.2

567

LAK cells rest	25.5	16.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.0	4.8
LAK cells IL-2	27.2	22.5	Liver cirrhosis	3.1	2.7
LAK cells IL-2+IL-12	27.2	19.3	Lupus kidney	2.1	1.7
LAK cells IL-2+IFN gamma	36.3	34.4	NCI-H292 none	30.1	18.9
LAK cells IL-2+ IL-18	35.1	29.7	NCI-H292 IL-4	33.9	34.6
LAK cells PMA/ionomycin	12.4	11.0	NCI-H292 IL-9	40.1	29.1
NK Cells IL-2 rest	20.0	15.0	NCI-H292 IL-13	16.2	14.2
Two Way MLR 3 day	24.0	16.7	NCI-H292 IFN gamma	16.6	18.4
Two Way MLR 5 day	12.9	10.1	HPAEC none	13.6	13.5
Two Way MLR 7 day	11.4	9.5	HPAEC TNF alpha + IL-1 beta	25.3	25.3
PBMC rest	13.7	10.5	Lung fibroblast none	11.4	14.2
PBMC PWM	69.3	66.4	Lung fibroblast TNF alpha + IL-1 beta	6.1	7.2
PBMC PHA-L	22.8	17.7	Lung fibroblast IL-4	28.5	29.1
Ramos (B cell) none	24.1	19.3	Lung fibroblast IL-9	23.0	23.3
Ramos (B cell) ionomycin	**100.0**	**100.0**	Lung fibroblast IL-13	20.6	18.9
B lymphocytes PWM	71.7	74.2	Lung fibroblast IFN gamma	39.0	32.5
B lymphocytes CD40L and IL-4	29.1	28.7	Dermal fibroblast CCD1070 rest	33.9	31.0
EOL-1 dbcAMP	12.1	10.5	Dermal fibroblast CCD1070 TNF alpha	76.8	62.0
EOL-1 dbcAMP PMA/ionomycin	14.5	10.9	Dermal fibroblast CCD1070 IL-1 beta	20.3	13.9
Dendritic cells none	13.2	14.8	Dermal fibroblast IFN gamma	14.2	9.5
Dendritic cells LPS	11.7	8.3	Dermal fibroblast IL-4	26.4	20.4
Dendritic cells	17.7	12.7	IBD Colitis 2	2.6	2.2

anti-CD40					
Monocytes rest	16.7	17.6	IBD Crohn's	2.0	1.9
Monocytes LPS	6.4	5.0	Colon	11.9	10.5
Macrophages rest	23.5	22.8	Lung	13.3	11.2
Macrophages LPS	9.9	7.1	Thymus	14.4	12.9
HUVEC none	20.6	17.9	Kidney	27.5	19.6
HUVEC starved	43.5	38.4			

CNS_neurodegeneration_v1.0 Summary: Ag3023/Ag3373 This panel does not show differential expression of the CG56804-04 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system. Ag5847 - This primer pair recognizes a splice variant of CG58564-01 designated CG58564-04. Expression of this variant is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3373 Highest expression of the CG56804-04 gene is seen in a prostate cancer cell line (CT=27). Overall, this gene is expressed at moderate levels in the cancer cell lines in this panel. A higher level of expression is observed in clusters of cell lines derived from prostate, brain, melanoma, colon, lung, breast and ovarian cancer when compared to expression in normal prostate, brain, colon, lung, breast and ovary. Thus, this gene could potentially be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be effective in the treatment of cancer in these tissues.

Among tissues with metabolic function, this gene product has moderate levels of expression in adipose, heart, skeletal muscle, adrenal, pituitary, thyroid and pancreas. Thus, this gene product may be a small molecule target for the treatment of endocrine and metabolic diseases, including obesity and Types 1 and 2 diabetes.

In addition, this gene appears to be differentially expressed in fetal (CT value = 29) vs adult liver (CT value =33) and may be useful for differentiation between the two sources of this tissue.

This gene is also expressed at moderate levels in all central nervous system samples present on this panel. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

General_screening_panel_v1.5 Summary: Ag5844 - This primer pair recognizes a splice variant of CG58564-01. Expression of this variant is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 1.3D Summary: Ag3023 The CG56804-04 gene is ubiquitously expressed among the samples on this panel, with highest expression in an ovarian cancer cell line (CT=28.8). Overall, the expression of this gene shows good agreement with panel 1.4. A higher level of expression is observed in prostate, brain, melanoma, colon, lung, pancreatic, breast and ovarian cancer cell lines than the normal prostate, brain, colon, lung, pancreas, breast and ovary. Thus, expression of this gene could be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be effective in the treatment of cancer in these tissues.

Among tissues with metabolic function, expression of this gene is widespread, as in the previous panel. Please see Panel 1.4 for discussion of utility of this gene in metabolic disease.

This gene represents a dual specificity phosphatase that is also expressed at low to moderate levels across the CNS. Dual-specificity phosphatases comprise a family of MAP kinase regulating enzymes, members of which are upregulated in brains subjected to insults such as ischemia and seizure activity. MAP kinases are known to regulate neurotrophic and neurotoxic pathways. Consequently, agents that modulate the activity of this gene may have utility in attenuating the apoptotic and neurodegenerative processes following brain insults.

Panel 4.1D Summary: Ag5844 - This primer pair recognizes a splice variant of CG58564-01. Expression of this variant is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag3023/Ag3373 The CG56804-04 gene is expressed at high to moderate levels in a wide range of cell types and tissues of significance in the immune response in health and disease. Highest expression of this gene is seen in ionomycin treated Ramos B cells (CT=26.83). Therefore, targeting of ghis gene product with a small molecule drug or antibody therapeutic may modulate the functions of cells of the immune system as well as resident tissue cells and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies,

inflammatory bowel disease, lupus erythematosus, and arthritis, including osteoarthritis and rheumatoid arthritis.

O. CG57819-01: RPGR-INTERACTING PROTEIN-1

Expression of gene CG57819-01 was assessed using the primer-probe set Ag3338, described in Table OA. Results of the RTQ-PCR runs are shown in Tables OB and OC.

Table OA. Probe Name Ag3338

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cccattcagcactgaaacag-3'	20	3021	426
Probe	TET-5'-tcctgtaaatgacaaagaatcctctgaaca-3'-TAMRA	30	3055	427
Reverse	5'-tgcttcactgacttcagaacct-3'	22	3085	428

Table OB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3338, Run 215773746	Tissue Name	Rel. Exp.(%) Ag3338, Run 215773746
Adipose	1.1	Renal ca. TK-10	0.8
Melanoma* Hs688(A).T	0.0	Bladder	1.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.2	Colon ca. SW480	0.4
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	**100.0**	Colon ca. HT29	0.5
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.2
Prostate Pool	1.0	Colon ca. CaCo-2	1.0
Placenta	0.0	Colon cancer tissue	0.9
Uterus Pool	0.0	Colon ca. SW1116	0.2
Ovarian ca. OVCAR-3	0.9	Colon ca. Colo-205	0.2
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0

Ovarian ca. OVCAR-4	1.2	Colon Pool	0.5
Ovarian ca. OVCAR-5	3.5	Small Intestine Pool	0.3
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.2
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.9	Fetal Heart	0.8
Breast ca. MCF-7	1.9	Heart Pool	1.1
Breast ca. MDA-MB-231	1.2	Lymph Node Pool	1.4
Breast ca. BT 549	0.2	Fetal Skeletal Muscle	0.2
Breast ca. T47D	6.7	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	1.4
Breast Pool	0.5	Thymus Pool	0.0
Trachea	0.9	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.2	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.4	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.8	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.5	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.1	CNS cancer (glio) SF-295	0.2
Lung ca. A549	1.5	Brain (Amygdala) Pool	0.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.6
Lung ca. NCI-H23	1.5	Brain (fetal)	0.9
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.7
Lung ca. HOP-62	3.0	Cerebral Cortex Pool	0.2
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.7
Liver	0.4	Brain (Thalamus) Pool	1.3
Fetal Liver	0.5	Brain (whole)	0.0
Liver ca. HepG2	0.2	Spinal Cord Pool	0.9
Kidney Pool	0.9	Adrenal Gland	0.0
Fetal Kidney	0.6	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0

Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	3.4
Renal ca. UO-31	0.0	Pancreas Pool	0.8

Table OC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3338, Run 165221737	Tissue Name	Rel. Exp.(%) Ag3338, Run 165221737
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	6.9
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	2.6
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	1.9
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	4.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	4.6	CCD1106	0.0

573

		(Keratinocytes) TNFalpha + IL-1beta	
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	2.4
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	4.5
LAK cells PMA/ionomycin	3.1	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	14.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	3.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	4.7	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	13.9	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	6.0	IBD Colitis 2	0.0
Monocytes rest	**100.0**	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	15.2
Macrophages rest	1.3	Lung	4.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	3.1
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag3338 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3338 - Expression of this gene is highest in testis (CT=29.4). Therefore, expression of this gene could be used to distinguish this sample from others on the panel.

There is also low expression in pancreatic cancer cell line CAPAN2, lung cancer cell line HOP-62, breast cancer cell line T47D, and ovarian cancer cell line OVCAR-5. Thus, expression of this gene could be used to differentiate these samples from other samples on this panel.

Panel 4D Summary: Ag3338 - Significant expression of this gene is seen only in resting monocytes (CT=32.3) Therefore, expression of this gene can be used to differentiate between this sample and others on this panel.

P. CG57789-01 and CG57789-02: RAS-LIKE PROTEIN RRP22-like

Expression of gene CG57789-01 and variant CG57789-02 was assessed using the primer-probe set Ag3333, described in Table PA. Results of the RTQ-PCR runs are shown in Tables PB, PC and PD.

Table PA. Probe Name Ag3333

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tcgactttccacccatcag-3'	19	181	429
Probe	TET-5'-cttccctgtcaatacgctccaggagt-3'-TAMRA	26	203	430
Reverse	5'-aggatgtaggcgtggacact-3'	20	258	431

Table PB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3333, Run 210146459	Tissue Name	Rel. Exp.(%) Ag3333, Run 210146459
AD 1 Hippo	22.2	Control (Path) 3 Temporal Ctx	7.5
AD 2 Hippo	18.8	Control (Path) 4 Temporal Ctx	21.6
AD 3 Hippo	17.9	AD 1 Occipital Ctx	29.7
AD 4 Hippo	8.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	**100.0**	AD 3 Occipital Ctx	15.8
AD 6 Hippo	42.9	AD 4 Occipital Ctx	24.7

575

Control 2 Hippo	25.9	AD 5 Occipital Ctx	90.1
Control 4 Hippo	12.1	AD 6 Occipital Ctx	16.3
Control (Path) 3 Hippo	13.4	Control 1 Occipital Ctx	4.2
AD 1 Temporal Ctx	21.3	Control 2 Occipital Ctx	74.7
AD 2 Temporal Ctx	29.1	Control 3 Occipital Ctx	14.5
AD 3 Temporal Ctx	13.3	Control 4 Occipital Ctx	4.5
AD 4 Temporal Ctx	15.8	Control (Path) 1 Occipital Ctx	47.3
AD 5 Inf Temporal Ctx	92.0	Control (Path) 2 Occipital Ctx	13.5
AD 5 Sup Temporal Ctx	43.2	Control (Path) 3 Occipital Ctx	4.1
AD 6 Inf Temporal Ctx	26.4	Control (Path) 4 Occipital Ctx	14.6
AD 6 Sup Temporal Ctx	31.6	Control 1 Parietal Ctx	7.6
Control 1 Temporal Ctx	5.8	Control 2 Parietal Ctx	39.2
Control 2 Temporal Ctx	51.8	Control 3 Parietal Ctx	21.9
Control 3 Temporal Ctx	14.5	Control (Path) 1 Parietal Ctx	56.3
Control 3 Temporal Ctx	8.1	Control (Path) 2 Parietal Ctx	20.2
Control (Path) 1 Temporal Ctx	39.2	Control (Path) 3 Parietal Ctx	6.2
Control (Path) 2 Temporal Ctx	40.9	Control (Path) 4 Parietal Ctx	24.5

Table PC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3333, Run 216516940	Tissue Name	Rel. Exp.(%) Ag3333, Run 216516940
Adipose	4.4	Renal ca. TK-10	40.1
Melanoma* Hs688(A).T	0.9	Bladder	5.0
Melanoma* Hs688(B).T	1.8	Gastric ca. (liver met.) NCI-N87	4.5
Melanoma* M14	2.7	Gastric ca. KATO III	20.0
Melanoma*	0.3	Colon ca. SW-948	0.0

576

LOXIMVI			
Melanoma* SK-MEL-5	0.9	Colon ca. SW480	**100.0**
Squamous cell carcinoma SCC-4	0.1	Colon ca.* (SW480 met) SW620	33.0
Testis Pool	2.1	Colon ca. HT29	5.0
Prostate ca.* (bone met) PC-3	2.4	Colon ca. HCT-116	0.1
Prostate Pool	0.5	Colon ca. CaCo-2	37.1
Placenta	5.7	Colon cancer tissue	2.3
Uterus Pool	0.3	Colon ca. SW1116	16.2
Ovarian ca. OVCAR-3	52.9	Colon ca. Colo-205	0.2
Ovarian ca. SK-OV-3	0.6	Colon ca. SW-48	0.2
Ovarian ca. OVCAR-4	17.9	Colon Pool	2.2
Ovarian ca. OVCAR-5	4.5	Small Intestine Pool	1.0
Ovarian ca. IGROV-1	0.9	Stomach Pool	0.9
Ovarian ca. OVCAR-8	15.4	Bone Marrow Pool	1.8
Ovary	4.2	Fetal Heart	10.9
Breast ca. MCF-7	0.7	Heart Pool	2.8
Breast ca. MDA-MB-231	0.4	Lymph Node Pool	4.4
Breast ca. BT 549	42.0	Fetal Skeletal Muscle	1.1
Breast ca. T47D	13.0	Skeletal Muscle Pool	46.7
Breast ca. MDA-N	0.1	Spleen Pool	0.0
Breast Pool	2.4	Thymus Pool	2.3
Trachea	2.4	CNS cancer (glio/astro) U87-MG	0.9
Lung	0.2	CNS cancer (glio/astro) U-118-MG	0.3
Fetal Lung	0.9	CNS cancer (neuro;met) SK-N-AS	69.7
Lung ca. NCI-N417	17.1	CNS cancer (astro) SF-539	2.2
Lung ca. LX-1	1.1	CNS cancer (astro) SNB-75	15.9
Lung ca. NCI-H146	14.5	CNS cancer (glio) SNB-19	0.6
Lung ca. SHP-77	37.6	CNS cancer (glio) SF-295	6.0

577

Lung ca. A549	0.4	Brain (Amygdala) Pool	28.5
Lung ca. NCI-H526	23.5	Brain (cerebellum)	29.1
Lung ca. NCI-H23	8.2	Brain (fetal)	21.3
Lung ca. NCI-H460	14.3	Brain (Hippocampus) Pool	27.7
Lung ca. HOP-62	1.7	Cerebral Cortex Pool	36.1
Lung ca. NCI-H522	86.5	Brain (Substantia nigra) Pool	40.1
Liver	1.6	Brain (Thalamus) Pool	37.6
Fetal Liver	0.7	Brain (whole)	59.5
Liver ca. HepG2	6.2	Spinal Cord Pool	12.3
Kidney Pool	3.8	Adrenal Gland	4.7
Fetal Kidney	7.4	Pituitary gland Pool	3.7
Renal ca. 786-0	0.2	Salivary Gland	48.0
Renal ca. A498	20.9	Thyroid (female)	1.1
Renal ca. ACHN	8.5	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	3.0	Pancreas Pool	4.0

Table PD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3333, Run 165084139	Tissue Name	Rel. Exp.(%) Ag3333, Run 165084139
Secondary Th1 act	0.8	HUVEC IL-1beta	0.0
Secondary Th2 act	3.0	HUVEC IFN gamma	0.5
Secondary Tr1 act	0.6	HUVEC TNF alpha + IFN gamma	0.8
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.5	HUVEC IL-11	0.3
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.6
Primary Th1 act	5.7	Lung Microvascular EC TNFalpha + IL-1beta	0.4
Primary Th2 act	9.8	Microvascular Dermal EC none	0.0
Primary Tr1 act	3.8	Microsvasular Dermal EC TNFalpha + IL-1beta	0.4
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	1.1
Primary Th2 rest	0.4	Small airway epithelium none	1.9
Primary Tr1 rest	0.6	Small airway epithelium	1.4

578

		TNFalpha + IL-1beta	
CD45RA CD4 lymphocyte act	4.1	Coronery artery SMC rest	1.7
CD45RO CD4 lymphocyte act	1.7	Coronery artery SMC TNFalpha + IL-1beta	1.2
CD8 lymphocyte act	1.4	Astrocytes rest	**100.0**
Secondary CD8 lymphocyte rest	7.4	Astrocytes TNFalpha + IL-1beta	59.9
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	2.0
CD4 lymphocyte none	0.8	KU-812 (Basophil) PMA/ionomycin	4.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.5	CCD1106 (Keratinocytes) none	12.5
LAK cells rest	0.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.2
LAK cells IL-2	0.3	Liver cirrhosis	0.9
LAK cells IL-2+IL-12	0.5	Lupus kidney	3.9
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	29.3
LAK cells IL-2+ IL-18	0.6	NCI-H292 IL-4	39.5
LAK cells PMA/ionomycin	0.3	NCI-H292 IL-9	23.3
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	21.9
Two Way MLR 3 day	0.8	NCI-H292 IFN gamma	14.5
Two Way MLR 5 day	0.9	HPAEC none	0.5
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	4.5
PBMC PWM	8.1	Lung fibroblast TNF alpha + IL-1 beta	2.2
PBMC PHA-L	11.6	Lung fibroblast IL-4	12.9
Ramos (B cell) none	0.0	Lung fibroblast IL-9	9.2
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	8.5
B lymphocytes PWM	15.4	Lung fibroblast IFN gamma	8.4
B lymphocytes CD40L and IL-4	2.1	Dermal fibroblast CCD1070 rest	40.6
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	20.9
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	19.3
Dendritic cells none	0.0	Dermal fibroblast IFN	1.8

		gamma	
Dendritic cells LPS	0.5	Dermal fibroblast IL-4	3.8
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	2.5
Monocytes LPS	0.0	Colon	4.2
Macrophages rest	0.0	Lung	9.1
Macrophages LPS	0.0	Thymus	11.3
HUVEC none	0.4	Kidney	2.6
HUVEC starved	0.6		

CNS_neurodegeneration_v1.0 Summary: This panel confirms the expression of this gene in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.4 Summary: Ag3333 This gene is expressed at moderate to low levels in many of the samples on this panel, with the highest expression in colon cancer cell line SW480 (CT=27.8). Expression is significantly lower in SW680, a cell line derived from a metastasis of the primary tumor represented by SW480. Thus, expression of this gene could be used to differentiate between these two cell lines and potentially between primary colon cancer and its metastases.

Based on expression in this panel, this gene may be involved in gastric, brain, colon, renal, lung, breast, ovarian and prostate cancer as well as melanomas. Thus, expression of this gene could be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition using antibodies or small molecule drugs might be of use in the treatment of these cancers.

This gene product is also expressed in adipose, pancreas, adrenal, thyroid, pituitary, skeletal muscle, heart, and liver. This widespread expression in tissues with metabolic function suggests that this gene product may be important for the pathogenesis, diagnosis, and/or treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes

This gene is expressed at low levels throughout the CNS, including in amygdala, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this

gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4D Summary: Ag3333 The CG57789-01 gene is expressed at moderate to low levels in several samples on this panel, with the highest expression in resting astrocytes (CT=28.4). Moderate expression of this gene is seen in treated and untreated dermal and lung fibroblasts and the airway epithelial tumor line NCI-H292 cells. Thus, the transcript or the protein it encodes may be involved in pathological and inflammatory skin and lung conditions, including psoriasis, asthma, allergy, emphysema, and COPD.

Q. CG57758-01 and CG57758-02: SODIUM/LITHIUM-DEPENDENT DICARBOXYLATE TRANSPORTER

Expression of gene CG57758-01, a splice variant of CG57758-02, and CG57758-02 was assessed using the primer-probe sets Ag3326 and Ag3692, described in Tables QA and QB. Results of the RTQ-PCR runs are shown in Tables QC, QD, QE and QF.

Table QA. Probe Name Ag3326

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ccatttactggtgcacagaagt-3'	22	149	432
Probe	TET-5'-atccctctggctgtcacctctctcat-3'-TAMRA	26	172	433
Reverse	5'-ggagtccagaatctggaagagt-3'	22	216	434

Table QB. Probe Name Ag3692

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ccatttactggtgcacagaagt-3'	22	149	435
Probe	TET-5'-atccctctggctgtcacctctctcat-3'-TAMRA	26	172	436
Reverse	5'-ggagtccagaatctggaagagt-3'	22	216	437

Table QC. CNS_neurodegeneration_v1.0

581

Tissue Name	Rel. Exp.(%) Ag3326, Run 210144197	Rel. Exp.(%) Ag3692, Run 211145262	Rel. Exp.(%) Ag3692, Run 224337942	Tissue Name	Rel. Exp.(%) Ag3326, Run 210144197	Rel. Exp.(%) Ag3692, Run 211145262	Rel. Exp.(%) Ag3692, Run 224337942
AD 1 Hippo	2.1	4.3	1.0	Control (Path) 3 Temporal Ctx	8.5	15.3	12.0
AD 2 Hippo	20.9	28.3	25.0	Control (Path) 4 Temporal Ctx	31.2	36.6	52.1
AD 3 Hippo	0.0	0.9	0.6	AD 1 Occipital Ctx	2.7	3.0	0.0
AD 4 Hippo	2.1	7.1	2.6	AD 2 Occipital Ctx (Missing)	0.0	0.0	0.0
AD 5 hippo	72.7	97.9	85.3	AD 3 Occipital Ctx	1.5	7.2	1.3
AD 6 Hippo	13.7	18.3	5.5	AD 4 Occipital Ctx	71.7	35.6	30.6
Control 2 Hippo	14.5	20.2	15.2	AD 5 Occipital Ctx	25.3	31.9	12.4
Control 4 Hippo	11.7	7.4	5.1	AD 6 Occipital Ctx	17.2	19.1	11.2
Control (Path) 3 Hippo	6.7	4.4	4.5	Control 1 Occipital Ctx	7.0	9.0	8.1
AD 1 Temporal Ctx	4.0	1.7	2.8	Control 2 Occipital Ctx	33.2	44.8	26.1
AD 2 Temporal Ctx	80.7	50.7	37.4	Control 3 Occipital Ctx	30.1	37.6	21.9
AD 3 Temporal Ctx	3.6	0.0	1.1	Control 4 Occipital	16.3	12.6	8.2

582

				Ctx			
AD 4 Temporal Ctx	19.5	30.6	15.2	Control (Path) 1 Occipital Ctx	42.0	55.9	52.9
AD 5 Inf Temporal Ctx	**100.0**	**100.0**	99.3	Control (Path) 2 Occipital Ctx	6.7	13.0	7.7
AD 5 SupTemporal Ctx	32.8	29.1	33.2	Control (Path) 3 Occipital Ctx	8.7	6.6	5.4
AD 6 Inf Temporal Ctx	27.7	21.3	26.6	Control (Path) 4 Occipital Ctx	8.1	9.0	7.4
AD 6 Sup Temporal Ctx	41.8	53.6	17.0	Control 1 Parietal Ctx	21.2	23.0	15.3
Control 1 Temporal Ctx	12.0	33.9	18.3	Control 2 Parietal Ctx	48.6	38.2	22.1
Control 2 Temporal Ctx	30.1	49.3	44.4	Control 3 Parietal Ctx	28.3	34.4	32.8
Control 3 Temporal Ctx	38.7	39.5	33.4	Control (Path) 1 Parietal Ctx	78.5	97.3	**100.0**
Control 4 Temporal Ctx	17.6	25.2	24.1	Control (Path) 2 Parietal Ctx	50.7	50.7	37.9
Control (Path) 1 Temporal Ctx	69.7	70.7	49.7	Control (Path) 3 Parietal Ctx	10.7	10.1	9.6
Control (Path) 2 Temporal Ctx	35.4	50.7	33.4	Control (Path) 4 Parietal Ctx	30.6	24.5	40.9

Table QD. General_screening_panel_v1.4

583

Tissue Name	Rel. Exp.(%) Ag3326, Run 215678613	Rel. Exp.(%) Ag3692, Run 217131191	Tissue Name	Rel. Exp.(%) Ag3326, Run 215678613	Rel. Exp.(%) Ag3692, Run 217131191
Adipose	0.0	0.0	Renal ca. TK-10	11.4	12.0
Melanoma* Hs688(A).T	0.0	0.0	Bladder	0.0	0.1
Melanoma* Hs688(B).T	0.1	0.0	Gastric ca. (liver met.) NCI-N87	0.0	0.0
Melanoma* M14	0.0	0.0	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMVI	0.0	0.0	Colon ca. SW-948	0.0	0.0
Melanoma* SK-MEL-5	0.0	0.0	Colon ca. SW480	0.0	0.0
Squamous cell carcinoma SCC-4	0.9	0.7	Colon ca.* (SW480 met) SW620	0.0	0.0
Testis Pool	0.1	0.2	Colon ca. HT29	0.0	0.0
Prostate ca.* (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	0.0	0.0
Prostate Pool	0.0	0.0	Colon ca. CaCo-2	0.0	0.0
Placenta	0.0	0.0	Colon cancer tissue	0.1	0.0
Uterus Pool	0.0	0.0	Colon ca. SW1116	0.0	0.0
Ovarian ca. OVCAR-3	0.0	0.0	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK-OV-3	0.0	0.0	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	0.1	0.0	Colon Pool	0.6	0.0
Ovarian ca. OVCAR-5	0.0	0.0	Small Intestine Pool	0.1	0.0
Ovarian ca. IGROV-1	0.0	0.0	Stomach Pool	0.0	0.0
Ovarian ca. OVCAR-8	2.8	2.2	Bone Marrow Pool	0.0	0.1
Ovary	0.7	0.6	Fetal Heart	0.0	0.0
Breast ca. MCF-7	0.0	0.0	Heart Pool	0.0	0.0
Breast ca. MDA-MB-231	0.0	0.0	Lymph Node Pool	0.1	0.0

Breast ca. BT 549	0.6	0.8	Fetal Skeletal Muscle	0.0	0.0
Breast ca. T47D	0.0	0.0	Skeletal Muscle Pool	0.0	0.0
Breast ca. MDA-N	0.0	0.0	Spleen Pool	0.4	0.2
Breast Pool	0.0	0.1	Thymus Pool	0.0	0.0
Trachea	0.2	0.1	CNS cancer (glio/astro) U87-MG	0.0	0.0
Lung	0.0	0.0	CNS cancer (glio/astro) U-118-MG	0.0	0.0
Fetal Lung	0.2	0.1	CNS cancer (neuro;met) SK-N-AS	0.0	0.0
Lung ca. NCI-N417	0.0	0.0	CNS cancer (astro) SF-539	0.0	0.0
Lung ca. LX-1	0.0	0.0	CNS cancer (astro) SNB-75	0.0	0.0
Lung ca. NCI-H146	0.0	0.0	CNS cancer (glio) SNB-19	0.0	0.0
Lung ca. SHP-77	0.0	0.0	CNS cancer (glio) SF-295	0.1	0.1
Lung ca. A549	0.0	0.1	Brain (Amygdala) Pool	0.4	0.4
Lung ca. NCI-H526	2.0	0.0	Brain (cerebellum)	1.4	1.0
Lung ca. NCI-H23	0.7	0.6	Brain (fetal)	0.7	0.4
Lung ca. NCI-H460	0.0	0.0	Brain (Hippocampus) Pool	0.5	0.7
Lung ca. HOP-62	0.1	0.2	Cerebral Cortex Pool	1.4	1.5
Lung ca. NCI-H522	0.0	0.0	Brain (Substantia nigra) Pool	1.4	1.4
Liver	28.7	24.1	Brain (Thalamus) Pool	1.1	0.9
Fetal Liver	100.0	100.0	Brain (whole)	4.1	3.7
Liver ca. HepG2	29.5	26.2	Spinal Cord Pool	0.1	0.2
Kidney Pool	0.0	0.0	Adrenal Gland	2.6	1.9
Fetal Kidney	0.1	0.1	Pituitary gland Pool	0.0	0.2
Renal ca.	0.0	0.0	Salivary Gland	40.9	35.1

786-0					
Renal ca. A498	0.0	0.0	Thyroid (female)	0.0	0.0
Renal ca. ACHN	0.0	0.0	Pancreatic ca. CAPAN2	0.5	0.8
Renal ca. UO-31	0.0	0.0	Pancreas Pool	0.0	0.0

Table QE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3692, Run 169987356	Tissue Name	Rel. Exp.(%) Ag3692, Run 169987356
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	11.3
Primary Tr1 act	4.2	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	28.5
Primary Th2 rest	0.0	Small airway epithelium none	5.7
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	3.9	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	3.6
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	4.3

2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	10.7
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	94.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	3.2	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	2.4
HUVEC none	0.0	Kidney	**100.0**
HUVEC starved	0.0		

Table QF. Panel 5 Islet

587

Tissue Name	Rel. Exp.(%) Ag3326, Run 242385365	Tissue Name	Rel. Exp.(%) Ag3326, Run 242385365
97457_Patient-02go_adipose	0.0	94709_Donor 2 AM - A_adipose	0.2
97476_Patient-07sk_skeletal muscle	0.0	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	0.0
99167_Bayer Patient 1	0.3	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	0.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	0.0	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	0.0	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	0.0	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-11ut_uterus	0.0	77138_Liver_HepG2untreated	**100.0**
97498_Patient-11pl_placenta	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.1	81735_Small Intestine	39.5
97501_Patient-12sk_skeletal muscle	0.3	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	0.0	82685_Small intestine_Duodenum	0.0
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	0.0

588

94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

CNS_neurodegeneration_v1.0 Summary: Ag3326/Ag3692 - Three experiments done with two primer pairs (same sequence) are in excellent agreement. This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's

5   diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.4 Summary: Ag3326/Ag3692 Two experiments with the smae probe and primer set produce results that are in excellent agreement. This gene is

10   highly expressed in fetal liver (CT=26.5-27.0) and moderately expressed in adult liver (CT=28.5-28.8) and liver cancer cell line HepG2 (CT=28.4-28.8). This result agrees with the results seen in Panel 5 (expression in HepG2 (CT=29.2). These results are in agreement with published data that show a novel sodium dicarboxylate transporter in brain, choroid plexus kidney, intestine and liver. Thus, expression of this gene could be used to

15   differentiate between these samples and other samples on this panel and as a marker for liver derived tissue.

This gene is expressed at low levels throughout the CNS, including in amygdala, substantia nigra, thalamus, cerebellum, and cerebral cortex. Therefore, this gene may play a role in central nervous system disorders such as Parkinson's disease, epilepsy, multiple

20   sclerosis, schizophrenia and depression.

Low but significant levels of expression are also seen in the adrenal gland. Thus, this gene product may also be involved in metabolic disorders of this gland, including adrenoleukodystrophy and congenital adrenal hyperplasia.

References:

589

1. Pajor AM, Gangula R, Yao X. Cloning and functional characterization of a high-affinity Na(+)/dicarboxylate cotransporter from mouse brain. Am J Physiol Cell Physiol 2001 May;280(5):C1215-23.

2. Chen XZ, Shayakul C, Berger UV, Tian W, Hediger MA. Characterization of a rat Na+-dicarboxylate cotransporter. J Biol Chem 1998 Aug 14;273(33):20972-81.

Panel 4.1D Summary: Ag3692 Significant expression of this gene is seen only in kidney and a liver cirrhosis sample (CTs=34.0). These results confirm that this gene is expressed in liver derived samples. The presence in the kidney is also in agreement with published results. Please see Panel 1.4. This gene product may be involved in maintaining or restoring normal function to the kidney during inflammation.

Panel 4D Summary: Ag3326 Results from one experiment are not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 5 Islet Summary: Ag3326 - The highest expression of this gene is in liver cancer cell line HepG2 (CT=29.2). There is also moderate expression in the small intestine (CT=30.5). These results compare well with previously published reports of sodium dicarboxylate transporter expression in mouse and rat (see discussion Panel 1.4).

R. CG57758-04 and CG57758-05: Sodium:sulfate symporter

Expression of gene CG57758-04 and CG57758-05, both splice variants of CG577584-01, was assessed using the primer-probe sets Ag3326, Ag3692 and Ag5818, described in Tables RA, RB and RC. Results of the RTQ-PCR runs are shown in Tables RD, RE, RF, RG and RH.

Table RA. Probe Name Ag3326

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ccatttactggtgcacagaagt-3'	22	138	438
Probe	TET-5'-atccctctggctgtcacctctctcat-3'-TAMRA	26	161	439
Reverse	5'-ggagtccagaatctggaagagt-3'	22	205	440

Table RB. Probe Name Ag3692

590

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ccatttactggtgcacagaagt-3'	22	138	441
Probe	TET-5'-atccctctggctgtcacctctctcat-3'-TAMRA	26	161	442
Reverse	5'-ggagtccagaatctggaagagt-3'	22	205	443

Table RC. Probe Name Ag5818

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ccatcaccttgatcttgtcc-3'	20	1341	444
Probe	TET-5'-ttatgactcctgttttcaccatggaggca-3'-TAMRA	29	1429	445
Reverse	5'-cagaagactccaattatgttca-3'	22	1458	446

Table RD. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3326, Run 210144197	Rel. Exp.(%) Ag3692, Run 211145262	Rel. Exp.(%) Ag3692, Run 224337942	Tissue Name	Rel. Exp.(%) Ag3326, Run 210144197	Rel. Exp.(%) Ag3692, Run 211145262	Rel. Exp.(%) Ag3692, Run 224337942
AD 1 Hippo	2.1	4.3	1.0	Control (Path) 3 Temporal Ctx	8.5	15.3	12.0
AD 2 Hippo	20.9	28.3	25.0	Control (Path) 4 Temporal Ctx	31.2	36.6	52.1
AD 3 Hippo	0.0	0.9	0.6	AD 1 Occipital Ctx	2.7	3.0	0.0
AD 4 Hippo	2.1	7.1	2.6	AD 2 Occipital Ctx (Missing)	0.0	0.0	0.0
AD 5 hippo	72.7	97.9	85.3	AD 3 Occipital Ctx	1.5	7.2	1.3
AD 6 Hippo	13.7	18.3	5.5	AD 4 Occipital Ctx	71.7	35.6	30.6
Control 2 Hippo	14.5	20.2	15.2	AD 5 Occipital	25.3	31.9	12.4

				Ctx			
Control 4 Hippo	11.7	7.4	5.1	AD 6 Occipital Ctx	17.2	19.1	11.2
Control (Path) 3 Hippo	6.7	4.4	4.5	Control 1 Occipital Ctx	7.0	9.0	8.1
AD 1 Temporal Ctx	4.0	1.7	2.8	Control 2 Occipital Ctx	33.2	44.8	26.1
AD 2 Temporal Ctx	80.7	50.7	37.4	Control 3 Occipital Ctx	30.1	37.6	21.9
AD 3 Temporal Ctx	3.6	0.0	1.1	Control 4 Occipital Ctx	16.3	12.6	8.2
AD 4 Temporal Ctx	19.5	30.6	15.2	Control (Path) 1 Occipital Ctx	42.0	55.9	52.9
AD 5 Inf Temporal Ctx	**100.0**	**100.0**	99.3	Control (Path) 2 Occipital Ctx	6.7	13.0	7.7
AD 5 SupTemporal Ctx	32.8	29.1	33.2	Control (Path) 3 Occipital Ctx	8.7	6.6	5.4
AD 6 Inf Temporal Ctx	27.7	21.3	26.6	Control (Path) 4 Occipital Ctx	8.1	9.0	7.4
AD 6 Sup Temporal Ctx	41.8	53.6	17.0	Control 1 Parietal Ctx	21.2	23.0	15.3
Control 1 Temporal Ctx	12.0	33.9	18.3	Control 2 Parietal Ctx	48.6	38.2	22.1
Control 2 Temporal Ctx	30.1	49.3	44.4	Control 3 Parietal Ctx	28.3	34.4	32.8

Control 3 Temporal Ctx	38.7	39.5	33.4	Control (Path) 1 Parietal Ctx	78.5	97.3	**100.0**
Control 4 Temporal Ctx	17.6	25.2	24.1	Control (Path) 2 Parietal Ctx	50.7	50.7	37.9
Control (Path) 1 Temporal Ctx	69.7	70.7	49.7	Control (Path) 3 Parietal Ctx	10.7	10.1	9.6
Control (Path) 2 Temporal Ctx	35.4	50.7	33.4	Control (Path) 4 Parietal Ctx	30.6	24.5	40.9

Table RE. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3326, Run 215678613	Rel. Exp.(%) Ag3692, Run 217131191	Tissue Name	Rel. Exp.(%) Ag3326, Run 215678613	Rel. Exp.(%) Ag3692, Run 217131191
Adipose	0.0	0.0	Renal ca. TK-10	11.4	12.0
Melanoma* Hs688(A).T	0.0	0.0	Bladder	0.0	0.1
Melanoma* Hs688(B).T	0.1	0.0	Gastric ca. (liver met.) NCI-N87	0.0	0.0
Melanoma* M14	0.0	0.0	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMVI	0.0	0.0	Colon ca. SW-948	0.0	0.0
Melanoma* SK-MEL-5	0.0	0.0	Colon ca. SW480	0.0	0.0
Squamous cell carcinoma SCC-4	0.9	0.7	Colon ca.* (SW480 met) SW620	0.0	0.0
Testis Pool	0.1	0.2	Colon ca. HT29	0.0	0.0
Prostate ca.* (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	0.0	0.0
Prostate Pool	0.0	0.0	Colon ca. CaCo-2	0.0	0.0
Placenta	0.0	0.0	Colon cancer tissue	0.1	0.0
Uterus Pool	0.0	0.0	Colon ca.	0.0	0.0

			SW1116		
Ovarian ca. OVCAR-3	0.0	0.0	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK-OV-3	0.0	0.0	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	0.1	0.0	Colon Pool	0.6	0.0
Ovarian ca. OVCAR-5	0.0	0.0	Small Intestine Pool	0.1	0.0
Ovarian ca. IGROV-1	0.0	0.0	Stomach Pool	0.0	0.0
Ovarian ca. OVCAR-8	2.8	2.2	Bone Marrow Pool	0.0	0.1
Ovary	0.7	0.6	Fetal Heart	0.0	0.0
Breast ca. MCF-7	0.0	0.0	Heart Pool	0.0	0.0
Breast ca. MDA-MB-231	0.0	0.0	Lymph Node Pool	0.1	0.0
Breast ca. BT 549	0.6	0.8	Fetal Skeletal Muscle	0.0	0.0
Breast ca. T47D	0.0	0.0	Skeletal Muscle Pool	0.0	0.0
Breast ca. MDA-N	0.0	0.0	Spleen Pool	0.4	0.2
Breast Pool	0.0	0.1	Thymus Pool	0.0	0.0
Trachea	0.2	0.1	CNS cancer (glio/astro) U87-MG	0.0	0.0
Lung	0.0	0.0	CNS cancer (glio/astro) U-118-MG	0.0	0.0
Fetal Lung	0.2	0.1	CNS cancer (neuro;met) SK-N-AS	0.0	0.0
Lung ca. NCI-N417	0.0	0.0	CNS cancer (astro) SF-539	0.0	0.0
Lung ca. LX-1	0.0	0.0	CNS cancer (astro) SNB-75	0.0	0.0
Lung ca. NCI-H146	0.0	0.0	CNS cancer (glio) SNB-19	0.0	0.0
Lung ca. SHP-77	0.0	0.0	CNS cancer (glio) SF-295	0.1	0.1
Lung ca. A549	0.0	0.1	Brain (Amygdala) Pool	0.4	0.4
Lung ca.	2.0	0.0	Brain	1.4	1.0

594

NCI-H526			(cerebellum)		
Lung ca. NCI-H23	0.7	0.6	Brain (fetal)	0.7	0.4
Lung ca. NCI-H460	0.0	0.0	Brain (Hippocampus) Pool	0.5	0.7
Lung ca. HOP-62	0.1	0.2	Cerebral Cortex Pool	1.4	1.5
Lung ca. NCI-H522	0.0	0.0	Brain (Substantia nigra) Pool	1.4	1.4
Liver	28.7	24.1	Brain (Thalamus) Pool	1.1	0.9
Fetal Liver	**100.0**	**100.0**	Brain (whole)	4.1	3.7
Liver ca. HepG2	29.5	26.2	Spinal Cord Pool	0.1	0.2
Kidney Pool	0.0	0.0	Adrenal Gland	2.6	1.9
Fetal Kidney	0.1	0.1	Pituitary gland Pool	0.0	0.2
Renal ca. 786-0	0.0	0.0	Salivary Gland	40.9	35.1
Renal ca. A498	0.0	0.0	Thyroid (female)	0.0	0.0
Renal ca. ACHN	0.0	0.0	Pancreatic ca. CAPAN2	0.5	0.8
Renal ca. UO-31	0.0	0.0	Pancreas Pool	0.0	0.0

Table RF. General_screening_panel_v1.5

Tissue Name	Rel. Exp.(%) Ag5818, Run 245382899	Tissue Name	Rel. Exp.(%) Ag5818, Run 245382899
Adipose	0.0	Renal ca. TK-10	13.4
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	1.4	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.5	Colon ca. HT29	0.0

Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.4
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	1.9	Bone Marrow Pool	0.0
Ovary	0.3	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.4	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.7
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.2	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.2	Brain (Amygdala) Pool	0.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	1.1
Lung ca. NCI-H23	1.5	Brain (fetal)	0.8
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.6

596

Lung ca. HOP-62	0.0	Cerebral Cortex Pool	1.7
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	1.2
Liver	40.3	Brain (Thalamus) Pool	1.3
Fetal Liver	**100.0**	Brain (whole)	5.6
Liver ca. HepG2	33.2	Spinal Cord Pool	0.3
Kidney Pool	0.0	Adrenal Gland	6.0
Fetal Kidney	0.0	Pituitary gland Pool	0.2
Renal ca. 786-0	0.0	Salivary Gland	67.4
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.7
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table RG. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3692, Run 169987356	Rel. Exp.(%) Ag5818, Run 246920287	Tissue Name	Rel. Exp.(%) Ag3692, Run 169987356	Rel. Exp.(%) Ag5818, Run 246920287
Secondary Th1 act	0.0	0.0	HUVEC IL-1beta	0.0	0.0
Secondary Th2 act	0.0	0.0	HUVEC IFN gamma	0.0	0.0
Secondary Tr1 act	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	0.0
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	0.0	0.0
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0	0.0
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	11.3	0.0
Primary Tr1 act	4.2	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0	0.0
Primary Th1 rest	0.0	0.0	Bronchial epithelium	28.5	0.0

597

			TNFalpha + IL1beta		
Primary Th2 rest	0.0	0.0	Small airway epithelium none	5.7	0.0
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0	0.0
CD45RA CD4 lymphocyte act	3.9	0.0	Coronery artery SMC rest	0.0	0.0
CD45RO CD4 lymphocyte act	0.0	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	0.0	0.0	Astrocytes rest	0.0	0.0
Secondary CD8 lymphocyte rest	0.0	0.0	Astrocytes TNFalpha + IL-1beta	0.0	0.0
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	3.6	24.3
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	4.3	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	0.0	CCD1106 (Keratinocytes) none	10.7	0.0
LAK cells rest	0.0	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0	0.0
LAK cells IL-2	0.0	0.0	Liver cirrhosis	94.0	27.5
LAK cells IL-2+IL-12	0.0	0.0	NCI-H292 none	0.0	0.0
LAK cells IL-2+IFN gamma	0.0	0.0	NCI-H292 IL-4	0.0	0.0
LAK cells IL-2+ IL-18	0.0	0.0	NCI-H292 IL-9	0.0	0.0
LAK cells PMA/ionomycin	0.0	0.0	NCI-H292 IL-13	0.0	0.0
NK Cells IL-2 rest	0.0	0.0	NCI-H292 IFN gamma	0.0	0.0
Two Way MLR 3 day	0.0	0.0	HPAEC none	0.0	0.0
Two Way MLR 5 day	3.2	0.0	HPAEC TNF alpha + IL-1 beta	0.0	0.0
Two Way MLR 7	0.0	0.0	Lung fibroblast	0.0	0.0

day			none		
PBMC rest	0.0	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0	0.0
PBMC PWM	0.0	0.0	Lung fibroblast IL-4	0.0	0.0
PBMC PHA-L	0.0	0.0	Lung fibroblast IL-9	0.0	0.0
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-13	0.0	0.0
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IFN gamma	0.0	0.0
B lymphocytes PWM	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0
B lymphocytes CD40L and IL-4	0.0	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0	0.0
EOL-1 dbcAMP	0.0	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	Dermal fibroblast IFN gamma	0.0	0.0
Dendritic cells none	0.0	0.0	Dermal fibroblast IL-4	0.0	0.0
Dendritic cells LPS	0.0	0.0	Dermal Fibroblasts rest	0.0	0.0
Dendritic cells anti-CD40	0.0	0.0	Neutrophils TNFa+LPS	0.0	0.0
Monocytes rest	0.0	0.0	Neutrophils rest	0.0	0.0
Monocytes LPS	0.0	0.0	Colon	0.0	0.0
Macrophages rest	0.0	0.0	Lung	0.0	0.0
Macrophages LPS	0.0	0.0	Thymus	2.4	0.0
HUVEC none	0.0	0.0	Kidney	100.0	100.0
HUVEC starved	0.0	0.0			

Table RH. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag3326, Run 242385365	Tissue Name	Rel. Exp.(%) Ag3326, Run 242385365
97457_Patient-02go_adipose	0.0	94709_Donor 2 AM - A_adipose	0.2
97476_Patient-07sk_skeletal muscle	0.0	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-	0.0	94711_Donor 2 AM - C_adipose	0.0

599

07ut_uterus			
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	0.0
99167_Bayer Patient 1	0.3	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	0.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	0.0	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	0.0	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	0.0	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-11ut_uterus	0.0	77138_Liver_HepG2untreated	**100.0**
97498_Patient-11pl_placenta	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.1	81735_Small Intestine	39.5
97501_Patient-12sk_skeletal muscle	0.3	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	0.0	82685_Small intestine_Duodenum	0.0
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	0.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

CNS_neurodegeneration_v1.0 Summary: Ag3326/Ag3692 - Three experiments done with two primer pairs (same sequence) are in excellent agreement. This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders. Ag5818 Results from one experiment are not included. The amp plot indicates that there were experimental difficulties with this run.

General_screening_panel_v1.4 Summary: Ag3326/Ag3692 Two experiments with the same probe and primer set produce results that are in excellent agreement. This gene is highly expressed in fetal liver (CT=26.5-27.0) and moderately expressed in adult liver (CT=28.5-28.8) and liver cancer cell line HepG2 (CT=28.4-28.8). This result agrees with the results seen in Panel 5 (expression in HepG2 (CT=29.2). These results are in agreement with published data that show a novel sodium dicarboxylate transporter in brain, choroid plexus kidney, intestine and liver. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker for liver derived tissue.

This gene is expressed at low levels throughout the CNS, including in amygdala, substantia nigra, thalamus, cerebellum, and cerebral cortex. Therefore, this gene may play a role in central nervous system disorders such as Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Low but significant levels of expression are also seen in the adrenal gland. Thus, this gene product may also be involved in metabolic disorders of this gland, including adrenoleukodystrophy and congenital adrenal hyperplasia.

References:

1. Pajor AM, Gangula R, Yao X. Cloning and functional characterization of a high-affinity Na(+)/dicarboxylate cotransporter from mouse brain. Am J Physiol Cell Physiol 2001 May;280(5):C1215-23.

2. Chen XZ, Shayakul C, Berger UV, Tian W, Hediger MA. Characterization of a rat Na+-dicarboxylate cotransporter. J Biol Chem 1998 Aug 14;273(33):20972-81.

General_screening_panel_v1.5 Summary: Ag5818 Results using this primer pair are in excellent agreement with the results seen in panel 1.4. See Panel 1.4 for discussion.

Panel 4.1D Summary: Ag3692 Significant expression of this gene is seen only in kidney and a liver cirrhosis sample (CTs=34.0). These results confirm that this gene is expressed in liver derived samples. The presence in the kidney is also in agreement with published results. Please see Panel 1.4. This gene product may be involved in maintaining or restoring normal function to the kidney during inflammation.

Panel 4D Summary: Ag3326 Results from one experiment are not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 5 Islet Summary: Ag3326 The highest expression of this gene is in liver cancer cell line HepG2 (CT=29.2). There is also moderate expression in the small intestine (CT=30.5). These results compare well with previously published reports of sodium dicarboxylate transporter expression in mouse and rat (see discussion Panel 1.4).

S. CG57732-01 and CG57732-02 and CG57732-03: CA2+/CALMODULIN-DEPENDENT PROTEIN KINASE IV KINASE

Expression of gene CG57732-01 and full length clones CG57732-02 and CG57732-03, was assessed using the primer-probe set Ag3317, described in Table SA. Results of the RTQ-PCR runs are shown in Tables SB, SC and SD. Please note CG57732-03 represents a splice variant of CG57732-01.

Table SA. Probe Name Ag3317

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggcctacaacgaaagtgaaga-3'	21	451	447
Probe	TET-5'-cagacactatgcaatgaaagtcctttcca-3'-TAMRA	29	472	448
Reverse	5'-ggaaagccatactgcttcagta-3'	22	510	449

Table SB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3317, Run 210144081	Tissue Name	Rel. Exp.(%) Ag3317, Run 210144081
AD 1 Hippo	10.7	Control (Path) 3	4.1

602

		Temporal Ctx	
AD 2 Hippo	23.7	Control (Path) 4 Temporal Ctx	42.6
AD 3 Hippo	4.5	AD 1 Occipital Ctx	12.5
AD 4 Hippo	7.5	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	97.9	AD 3 Occipital Ctx	5.4
AD 6 Hippo	25.7	AD 4 Occipital Ctx	18.4
Control 2 Hippo	24.8	AD 5 Occipital Ctx	21.8
Control 4 Hippo	4.3	AD 6 Occipital Ctx	58.6
Control (Path) 3 Hippo	2.8	Control 1 Occipital Ctx	1.5
AD 1 Temporal Ctx	10.4	Control 2 Occipital Ctx	94.0
AD 2 Temporal Ctx	35.8	Control 3 Occipital Ctx	21.5
AD 3 Temporal Ctx	5.8	Control 4 Occipital Ctx	2.6
AD 4 Temporal Ctx	23.2	Control (Path) 1 Occipital Ctx	**100.0**
AD 5 Inf Temporal Ctx	88.9	Control (Path) 2 Occipital Ctx	13.8
AD 5 SupTemporal Ctx	26.6	Control (Path) 3 Occipital Ctx	0.9
AD 6 Inf Temporal Ctx	39.5	Control (Path) 4 Occipital Ctx	19.6
AD 6 Sup Temporal Ctx	47.3	Control 1 Parietal Ctx	4.9
Control 1 Temporal Ctx	4.4	Control 2 Parietal Ctx	33.0
Control 2 Temporal Ctx	63.3	Control 3 Parietal Ctx	27.4
Control 3 Temporal Ctx	20.4	Control (Path) 1 Parietal Ctx	95.9
Control 4 Temporal Ctx	8.7	Control (Path) 2 Parietal Ctx	24.5
Control (Path) 1 Temporal Ctx	77.4	Control (Path) 3 Parietal Ctx	2.0
Control (Path) 2 Temporal Ctx	38.7	Control (Path) 4 Parietal Ctx	51.8

Table SC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3317, Run	Tissue Name	Rel. Exp.(%) Ag3317, Run

	215678602		215678602
Adipose	2.4	Renal ca. TK-10	14.2
Melanoma* Hs688(A).T	6.2	Bladder	10.5
Melanoma* Hs688(B).T	7.9	Gastric ca. (liver met.) NCI-N87	22.2
Melanoma* M14	18.2	Gastric ca. KATO III	23.0
Melanoma* LOXIMVI	9.4	Colon ca. SW-948	11.1
Melanoma* SK-MEL-5	9.8	Colon ca. SW480	20.9
Squamous cell carcinoma SCC-4	1.6	Colon ca.* (SW480 met) SW620	21.6
Testis Pool	13.1	Colon ca. HT29	11.3
Prostate ca.* (bone met) PC-3	6.4	Colon ca. HCT-116	27.0
Prostate Pool	3.1	Colon ca. CaCo-2	1.6
Placenta	1.8	Colon cancer tissue	11.3
Uterus Pool	3.9	Colon ca. SW1116	9.7
Ovarian ca. OVCAR-3	11.6	Colon ca. Colo-205	1.7
Ovarian ca. SK-OV-3	18.7	Colon ca. SW-48	8.8
Ovarian ca. OVCAR-4	3.4	Colon Pool	17.1
Ovarian ca. OVCAR-5	17.2	Small Intestine Pool	21.2
Ovarian ca. IGROV-1	6.2	Stomach Pool	5.3
Ovarian ca. OVCAR-8	4.7	Bone Marrow Pool	5.1
Ovary	2.9	Fetal Heart	6.8
Breast ca. MCF-7	6.1	Heart Pool	5.4
Breast ca. MDA-MB-231	20.3	Lymph Node Pool	13.4
Breast ca. BT 549	7.4	Fetal Skeletal Muscle	2.6
Breast ca. T47D	37.9	Skeletal Muscle Pool	2.3
Breast ca. MDA-N	9.0	Spleen Pool	2.8
Breast Pool	12.0	Thymus Pool	9.0
Trachea	17.2	CNS cancer (glio/astro) U87-MG	66.4
Lung	0.7	CNS cancer (glio/astro) U-118-MG	53.2
Fetal Lung	6.0	CNS cancer	4.6

		(neuro;met) SK-N-AS	
Lung ca. NCI-N417	16.5	CNS cancer (astro) SF-539	17.2
Lung ca. LX-1	20.9	CNS cancer (astro) SNB-75	21.5
Lung ca. NCI-H146	7.0	CNS cancer (glio) SNB-19	5.1
Lung ca. SHP-77	23.0	CNS cancer (glio) SF-295	12.2
Lung ca. A549	23.7	Brain (Amygdala) Pool	46.3
Lung ca. NCI-H526	4.4	Brain (cerebellum)	92.7
Lung ca. NCI-H23	5.8	Brain (fetal)	25.7
Lung ca. NCI-H460	10.3	Brain (Hippocampus) Pool	42.9
Lung ca. HOP-62	7.0	Cerebral Cortex Pool	**100.0**
Lung ca. NCI-H522	2.9	Brain (Substantia nigra) Pool	76.3
Liver	0.1	Brain (Thalamus) Pool	63.7
Fetal Liver	1.3	Brain (whole)	56.6
Liver ca. HepG2	1.4	Spinal Cord Pool	9.3
Kidney Pool	26.2	Adrenal Gland	16.2
Fetal Kidney	3.5	Pituitary gland Pool	16.4
Renal ca. 786-0	26.4	Salivary Gland	13.4
Renal ca. A498	14.2	Thyroid (female)	6.3
Renal ca. ACHN	33.2	Pancreatic ca. CAPAN2	2.6
Renal ca. UO-31	4.3	Pancreas Pool	19.6

Table SD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3317, Run 164683049	Tissue Name	Rel. Exp.(%) Ag3317, Run 164683049
Secondary Th1 act	21.6	HUVEC IL-1beta	3.8
Secondary Th2 act	23.2	HUVEC IFN gamma	12.5
Secondary Tr1 act	22.8	HUVEC TNF alpha + IFN gamma	2.9
Secondary Th1 rest	12.7	HUVEC TNF alpha + IL4	9.0
Secondary Th2 rest	9.3	HUVEC IL-11	4.0
Secondary Tr1 rest	33.7	Lung Microvascular EC none	24.3
Primary Th1 act	44.1	Lung Microvascular EC TNFalpha + IL-1beta	11.3

Primary Th2 act	49.3	Microvascular Dermal EC none	41.5
Primary Tr1 act	74.2	Microsvasular Dermal EC TNFalpha + IL-1beta	17.2
Primary Th1 rest	38.2	Bronchial epithelium TNFalpha + IL1beta	31.2
Primary Th2 rest	44.4	Small airway epithelium none	8.0
Primary Tr1 rest	50.0	Small airway epithelium TNFalpha + IL-1beta	11.1
CD45RA CD4 lymphocyte act	41.2	Coronery artery SMC rest	20.6
CD45RO CD4 lymphocyte act	25.0	Coronery artery SMC TNFalpha + IL-1beta	19.6
CD8 lymphocyte act	17.8	Astrocytes rest	14.7
Secondary CD8 lymphocyte rest	21.8	Astrocytes TNFalpha + IL-1beta	11.0
Secondary CD8 lymphocyte act	7.4	KU-812 (Basophil) rest	2.1
CD4 lymphocyte none	21.8	KU-812 (Basophil) PMA/ionomycin	12.9
2ry Th1/Th2/Tr1_anti-CD95 CH11	5.8	CCD1106 (Keratinocytes) none	30.6
LAK cells rest	51.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	23.7
LAK cells IL-2	7.0	Liver cirrhosis	0.8
LAK cells IL-2+IL-12	25.5	Lupus kidney	2.2
LAK cells IL-2+IFN gamma	35.4	NCI-H292 none	33.7
LAK cells IL-2+ IL-18	28.7	NCI-H292 IL-4	43.5
LAK cells PMA/ionomycin	20.6	NCI-H292 IL-9	36.3
NK Cells IL-2 rest	13.5	NCI-H292 IL-13	35.6
Two Way MLR 3 day	33.0	NCI-H292 IFN gamma	24.3
Two Way MLR 5 day	9.6	HPAEC none	22.8
Two Way MLR 7 day	10.0	HPAEC TNF alpha + IL-1 beta	8.3
PBMC rest	12.0	Lung fibroblast none	11.8
PBMC PWM	24.7	Lung fibroblast TNF alpha + IL-1 beta	1.2
PBMC PHA-L	32.5	Lung fibroblast IL-4	19.2
Ramos (B cell) none	1.5	Lung fibroblast IL-9	12.1
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	14.8

606

B lymphocytes PWM	41.2	Lung fibroblast IFN gamma	17.2
B lymphocytes CD40L and IL-4	14.5	Dermal fibroblast CCD1070 rest	**100.0**
EOL-1 dbcAMP	20.0	Dermal fibroblast CCD1070 TNF alpha	57.8
EOL-1 dbcAMP PMA/ionomycin	60.3	Dermal fibroblast CCD1070 IL-1 beta	14.2
Dendritic cells none	55.5	Dermal fibroblast IFN gamma	24.1
Dendritic cells LPS	26.1	Dermal fibroblast IL-4	39.0
Dendritic cells anti-CD40	74.7	IBD Colitis 2	1.6
Monocytes rest	48.0	IBD Crohn's	2.7
Monocytes LPS	15.4	Colon	19.1
Macrophages rest	98.6	Lung	14.4
Macrophages LPS	5.6	Thymus	10.5
HUVEC none	27.9	Kidney	**100.0**
HUVEC starved	27.0		

CNS_neurodegeneration_v1.0 Summary: Ag3317 - This panel does not show differential expression of this gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.4 for discussion of utility of this gene in the central nervous system.

General_screening_panel_v1.4 Summary: Ag3317 - There is low to moderate expression this gene across all samples on this panel. This gene is expressed at moderate levels throughout the CNS, including in amygdala, substantia nigra, thalamus, cerebellum, and cerebral cortex. Highest expression is observed in the cerebral cortex (CT=29.0). This gene encodes a calmodulin-dependent protein kinase IV homolog, which is known to play a role in Ca2+ signaling in the CNS that controls neuronal growth, differentiation, and plasticity. Mice deficient in calmodulin-dependent protein kinase IV were found to have cerebellar defects. Therefore, this gene may play a role in central nervous system disorders such as Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

This gene product is also expressed in adipose, pancreas, adrenal, thyroid, pituitary, skeletal muscle, heart, and liver. This widespread expression in tissues with metabolic function suggests that this gene product may be important for the pathogenesis, diagnosis,

and/or treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

Based on expression in this panel, this gene may be also be involved in gastric, pancreatic, brain, colon, renal, lung, breast, ovarian and prostate cancer as well as melanomas. Thus, expression of this gene could be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition using antibodies or small molecule drugs might be of use in the treatment of these cancers.

References:

1. Okuno S, Kitani T, Fujisawa H. Evidence for the existence of Ca2+/calmodulin-dependent protein kinase IV kinase isoforms in rat brain. J Biochem (Tokyo) 1996 Jun;119(6):1176-81.

2. Ribar TJ, Rodriguiz RM, Khiroug L, Wetsel WC, Augustine GJ, Means AR. Cerebellar defects in Ca2+/calmodulin kinase IV-deficient mice. J Neurosci 2000 Nov 15;20(22):RC107.

Panel 4D Summary: Ag3317 - This gene was found to have low expression across almost all the samples on this panel, with the highest level of expression seen in kidney and resting dermal fibroblasts (CTs=32). Expression of Ca2+/calmodulin-dependent kinase type IV in thymocytes has been found in mice, where it plays a role in Ca2+-dependent gene transcription.

Reference

1. Raman V, Blaeser F, Ho N, Engle DL, Williams CB, Chatila TA. Requirement for Ca2+/calmodulin-dependent kinase type IV/Gr in setting the thymocyte selection threshold. J Immunol 2001 Dec 1;167(11):6270-8.

T. CG57709-01: Novel mitochondrial protein

Expression of gene CG57709-01 was assessed using the primer-probe set Ag3323, described in Table TA. Results of the RTQ-PCR runs are shown in Tables TB, TC and TD.

Table TA. Probe Name Ag3323

608

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atgtgcagaggatacgcatg-3'	20	589	450
Probe	TET-5'-tgcaaaacaggaagacaaaggaaggg-3'-TAMRA	26	626	451
Reverse	5'-tggttctggcattctagacg-3'	20	665	452

Table TB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3323, Run 210144152	Tissue Name	Rel. Exp.(%) Ag3323, Run 210144152
AD 1 Hippo	22.5	Control (Path) 3 Temporal Ctx	5.2
AD 2 Hippo	29.5	Control (Path) 4 Temporal Ctx	32.5
AD 3 Hippo	6.9	AD 1 Occipital Ctx	18.6
AD 4 Hippo	7.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	82.4	AD 3 Occipital Ctx	7.6
AD 6 Hippo	66.4	AD 4 Occipital Ctx	17.8
Control 2 Hippo	27.5	AD 5 Occipital Ctx	30.8
Control 4 Hippo	11.9	AD 6 Occipital Ctx	48.6
Control (Path) 3 Hippo	8.4	Control 1 Occipital Ctx	4.0
AD 1 Temporal Ctx	18.6	Control 2 Occipital Ctx	58.2
AD 2 Temporal Ctx	30.6	Control 3 Occipital Ctx	14.2
AD 3 Temporal Ctx	7.6	Control 4 Occipital Ctx	6.6
AD 4 Temporal Ctx	21.5	Control (Path) 1 Occipital Ctx	70.7
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	12.6
AD 5 SupTemporal Ctx	42.6	Control (Path) 3 Occipital Ctx	2.4
AD 6 Inf Temporal Ctx	48.6	Control (Path) 4 Occipital Ctx	14.6
AD 6 Sup Temporal Ctx	42.0	Control 1 Parietal Ctx	6.5

Control 1 Temporal Ctx	6.3	Control 2 Parietal Ctx	48.0
Control 2 Temporal Ctx	39.0	Control 3 Parietal Ctx	19.6
Control 3 Temporal Ctx	13.1	Control (Path) 1 Parietal Ctx	61.1
Control 4 Temporal Ctx	8.9	Control (Path) 2 Parietal Ctx	19.3
Control (Path) 1 Temporal Ctx	53.6	Control (Path) 3 Parietal Ctx	3.8
Control (Path) 2 Temporal Ctx	34.2	Control (Path) 4 Parietal Ctx	42.6

Table TC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3323, Run 165678151	Tissue Name	Rel. Exp.(%) Ag3323, Run 165678151
Liver adenocarcinoma	25.0	Kidney (fetal)	6.5
Pancreas	12.8	Renal ca. 786-0	14.3
Pancreatic ca. CAPAN 2	24.5	Renal ca. A498	34.2
Adrenal gland	12.2	Renal ca. RXF 393	14.2
Thyroid	6.9	Renal ca. ACHN	12.9
Salivary gland	14.0	Renal ca. UO-31	48.6
Pituitary gland	10.1	Renal ca. TK-10	7.2
Brain (fetal)	13.7	Liver	20.2
Brain (whole)	29.7	Liver (fetal)	22.1
Brain (amygdala)	21.3	Liver ca. (hepatoblast) HepG2	21.3
Brain (cerebellum)	24.7	Lung	6.7
Brain (hippocampus)	25.7	Lung (fetal)	14.8
Brain (substantia nigra)	20.0	Lung ca. (small cell) LX-1	39.8
Brain (thalamus)	27.2	Lung ca. (small cell) NCI-H69	25.0
Cerebral Cortex	33.0	Lung ca. (s.cell var.) SHP-77	42.3
Spinal cord	16.5	Lung ca. (large cell)NCI-H460	25.7
glio/astro U87-MG	8.9	Lung ca. (non-sm. cell) A549	12.0
glio/astro U-118-MG	**100.0**	Lung ca. (non-s.cell) NCI-H23	9.1

610

astrocytoma SW1783	14.6	Lung ca. (non-s.cell) HOP-62	9.5
neuro*; met SK-N-AS	43.2	Lung ca. (non-s.cl) NCI-H522	10.7
astrocytoma SF-539	13.9	Lung ca. (squam.) SW 900	12.4
astrocytoma SNB-75	29.7	Lung ca. (squam.) NCI-H596	59.0
glioma SNB-19	13.5	Mammary gland	10.6
glioma U251	43.8	Breast ca.* (pl.ef) MCF-7	46.3
glioma SF-295	17.7	Breast ca.* (pl.ef) MDA-MB-231	31.6
Heart (fetal)	22.7	Breast ca.* (pl.ef) T47D	15.1
Heart	14.5	Breast ca. BT-549	54.0
Skeletal muscle (fetal)	6.8	Breast ca. MDA-N	11.5
Skeletal muscle	55.5	Ovary	8.7
Bone marrow	10.7	Ovarian ca. OVCAR-3	26.2
Thymus	5.5	Ovarian ca. OVCAR-4	21.6
Spleen	13.3	Ovarian ca. OVCAR-5	20.9
Lymph node	24.8	Ovarian ca. OVCAR-8	12.6
Colorectal	8.8	Ovarian ca. IGROV-1	4.4
Stomach	15.1	Ovarian ca.* (ascites) SK-OV-3	23.5
Small intestine	28.3	Uterus	14.3
Colon ca. SW480	27.5	Placenta	6.9
Colon ca.* SW620(SW480 met)	17.6	Prostate	9.5
Colon ca. HT29	14.6	Prostate ca.* (bone met)PC-3	17.7
Colon ca. HCT-116	43.2	Testis	10.1
Colon ca. CaCo-2	12.7	Melanoma Hs688(A).T	7.6
Colon ca. tissue(ODO3866)	22.5	Melanoma* (met) Hs688(B).T	6.6
Colon ca. HCC-2998	25.2	Melanoma UACC-62	19.6
Gastric ca.* (liver met) NCI-N87	29.5	Melanoma M14	39.2

Bladder	6.1	Melanoma LOX IMVI	13.4
Trachea	13.2	Melanoma* (met) SK-MEL-5	21.2
Kidney	15.6	Adipose	6.5

Table TD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3323, Run 165296416	Tissue Name	Rel. Exp.(%) Ag3323, Run 165296416
Secondary Th1 act	32.3	HUVEC IL-1beta	3.8
Secondary Th2 act	22.8	HUVEC IFN gamma	12.0
Secondary Tr1 act	29.9	HUVEC TNF alpha + IFN gamma	8.1
Secondary Th1 rest	3.8	HUVEC TNF alpha + IL4	11.1
Secondary Th2 rest	4.3	HUVEC IL-11	8.4
Secondary Tr1 rest	6.0	Lung Microvascular EC none	7.6
Primary Th1 act	33.0	Lung Microvascular EC TNFalpha + IL-1beta	6.9
Primary Th2 act	25.0	Microvascular Dermal EC none	14.7
Primary Tr1 act	40.1	Microsvasular Dermal EC TNFalpha + IL-1beta	7.6
Primary Th1 rest	17.8	Bronchial epithelium TNFalpha + IL1beta	17.3
Primary Th2 rest	11.6	Small airway epithelium none	6.6
Primary Tr1 rest	15.0	Small airway epithelium TNFalpha + IL-1beta	18.4
CD45RA CD4 lymphocyte act	15.0	Coronery artery SMC rest	9.9
CD45RO CD4 lymphocyte act	24.7	Coronery artery SMC TNFalpha + IL-1beta	6.5
CD8 lymphocyte act	19.3	Astrocytes rest	5.1
Secondary CD8 lymphocyte rest	22.7	Astrocytes TNFalpha + IL-1beta	3.9
Secondary CD8 lymphocyte act	12.9	KU-812 (Basophil) rest	14.0
CD4 lymphocyte none	2.9	KU-812 (Basophil) PMA/ionomycin	22.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	5.4	CCD1106 (Keratinocytes) none	16.0

612

LAK cells rest	7.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	8.1
LAK cells IL-2	17.2	Liver cirrhosis	1.7
LAK cells IL-2+IL-12	15.1	Lupus kidney	1.0
LAK cells IL-2+IFN gamma	27.9	NCI-H292 none	30.1
LAK cells IL-2+ IL-18	17.7	NCI-H292 IL-4	49.0
LAK cells PMA/ionomycin	1.9	NCI-H292 IL-9	33.2
NK Cells IL-2 rest	8.4	NCI-H292 IL-13	26.2
Two Way MLR 3 day	9.9	NCI-H292 IFN gamma	26.6
Two Way MLR 5 day	18.4	HPAEC none	11.7
Two Way MLR 7 day	8.9	HPAEC TNF alpha + IL-1 beta	7.5
PBMC rest	3.8	Lung fibroblast none	8.0
PBMC PWM	50.3	Lung fibroblast TNF alpha + IL-1 beta	5.5
PBMC PHA-L	29.3	Lung fibroblast IL-4	19.1
Ramos (B cell) none	33.9	Lung fibroblast IL-9	15.3
Ramos (B cell) ionomycin	83.5	Lung fibroblast IL-13	11.4
B lymphocytes PWM	**100.0**	Lung fibroblast IFN gamma	16.5
B lymphocytes CD40L and IL-4	22.4	Dermal fibroblast CCD1070 rest	28.9
EOL-1 dbcAMP	10.5	Dermal fibroblast CCD1070 TNF alpha	31.2
EOL-1 dbcAMP PMA/ionomycin	3.7	Dermal fibroblast CCD1070 IL-1 beta	11.3
Dendritic cells none	9.9	Dermal fibroblast IFN gamma	5.2
Dendritic cells LPS	6.3	Dermal fibroblast IL-4	12.3
Dendritic cells anti-CD40	7.3	IBD Colitis 2	0.7
Monocytes rest	7.0	IBD Crohn's	1.0
Monocytes LPS	1.4	Colon	8.8
Macrophages rest	13.5	Lung	5.8
Macrophages LPS	3.5	Thymus	17.3
HUVEC none	11.9	Kidney	12.3
HUVEC starved	24.8		

CNS_neurodegeneration_v1.0 Summary: Ag3323 This panel does not show differential expression of the CG57709-01 gene in Alzheimer's disease. However, this expression

613

profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3323 - This gene is expressed at moderate levels in all samples on this panel, with highest expression in a brain cancer cell line. Expression is also seen in all the cancer cell lines on this panel. Thus, expression of this gene could be used to differentiate between this brain cancer cell line sample and other samples on this panel and as a marker for brain cancer.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic and that disregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This molecule is also expressed at moderate to low levels in the CNS and may be a small molecule target for the treatment of neurologic diseases such as Alzheimer's disease, Parkinson's disease, epilepsy, schizophrenia, stroke and multiple sclerosis.

Panel 4D Summary: Ag3323 - This gene is expressed at high to moderate levels in all samples on this panel, with highest expression in B lymphocytes stimulated with polkweed mitogen (CT=24.5). In addition, this gene is expressed at higher levels in ionomycin-activated Ramos B lymphocytes. The highl levels of expression in activated B lymphocytes suggests that therapies that antagonize the function of this gene product may reduce or eliminate the symptoms in patients with autoimmune and inflammatory diseases in which B cells play a part in the initiation or progression of the disease process, such as lupus erythematosus, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, or psoriasis.

U. CG57700-01: HYDROXYACYLGLUTATHIONE HYDROLASE (GLYOXALASE II)

Expression of gene CG57700-01 was assessed using the primer-probe set Ag3311, described in Table UA. Results of the RTQ-PCR runs are shown in Table UB.

Table UA. Probe Name Ag3311

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-acgcttagcaacctggagtt-3'	20	536	453
Probe	TET-5'-accacgtgagagccaagctgtcct-3'-TAMRA	24	582	454
Reverse	5'-gtcatcctcatccctcttctg-3'	21	611	456

Table UB. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3311, Run 164682845	Tissue Name	Rel. Exp.(%) Ag3311, Run 164682845
Secondary Th1 act	10.2	HUVEC IL-1beta	0.0
Secondary Th2 act	3.8	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	1.6	Microvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	5.1
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-	0.0	CCD1106	4.2

CD95 CH11		(Keratinocytes) none	
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	2.7
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	4.5	NCI-H292 IFN gamma	1.9
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	3.7	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	14.1
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	4.3
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	1.6	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	2.1	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	3.0
Dendritic cells anti-CD40	2.5	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	**100.0**
Macrophages rest	0.0	Lung	29.3
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	2.4
HUVEC starved	0.0		

AI_comprehensive panel_v1.0 Summary: Ag3311 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

CNS_neurodegeneration_v1.0 Summary: Ag3311 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

5   **General_screening_panel_v1.4 Summary:** Ag3311 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag3311 - Significant expression of this gene is seen only in colon (CT=33.9). Therefore, expression of this gene can be used to distinguish between this sample and the others on the panel and between healthy and inflammed bowel. Since

10   expression is not detectable in samples derived from Crohn's and colitis patients, therapeutic modulation of the expression or function of this gene may be useful in the treatment of inflammatory bowel disease.

V. CG58553-01: vasolpressin receptor

Expression of gene CG58553-01 was assessed using the primer-probe set Ag3372,

15   described in Table VA. Results of the RTQ-PCR runs are shown in Tables VB and VC.

Table VA. Probe Name Ag3372

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cggatctggtcatcacaca-3'	19	1983	457
Probe	TET-5'-ccacccacaacctcccaaggaact-3'-TAMRA	24	2017	458
Reverse	5'-agcctcagaaggtcgagatg-3'	20	2041	459

Table VB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3372, Run 165524269	Tissue Name	Rel. Exp.(%) Ag3372, Run 165524269
Liver adenocarcinoma	0.7	Kidney (fetal)	0.0
Pancreas	0.2	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.1
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.1	Renal ca. ACHN	0.0

Salivary gland	0.1	Renal ca. UO-31	0.0
Pituitary gland	0.2	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	2.1
Brain (whole)	0.3	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.2
Brain (cerebellum)	0.1	Lung	2.4
Brain (hippocampus)	0.5	Lung (fetal)	0.2
Brain (substantia nigra)	0.2	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.1
Spinal cord	1.0	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.1
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.6
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.1
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.1	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.4	Mammary gland	0.7
glioma U251	0.2	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.1
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	1.6	Ovarian ca. OVCAR-3	0.2
Thymus	1.7	Ovarian ca. OVCAR-4	0.0
Spleen	2.8	Ovarian ca. OVCAR-5	0.2
Lymph node	5.5	Ovarian ca.	0.2

618

		OVCAR-8	
Colorectal	0.2	Ovarian ca. IGROV-1	0.0
Stomach	1.2	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	**100.0**	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.8
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.1
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	1.4
Colon ca. CaCo-2	0.3	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.7	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	3.8	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	1.0	Melanoma M14	0.2
Bladder	0.0	Melanoma LOX IMVI	0.2
Trachea	0.1	Melanoma* (met) SK-MEL-5	0.4
Kidney	0.6	Adipose	1.3

Table VC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3372, Run 165296616	Tissue Name	Rel. Exp.(%) Ag3372, Run 165296616
Secondary Th1 act	1.4	HUVEC IL-1beta	0.0
Secondary Th2 act	1.4	HUVEC IFN gamma	0.0
Secondary Tr1 act	2.9	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	5.4	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	6.4	HUVEC IL-11	0.0
Secondary Tr1 rest	12.0	Lung Microvascular EC none	0.0
Primary Th1 act	11.4	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	18.9	Microvascular Dermal EC none	0.0
Primary Tr1 act	27.0	Microsvasular Dermal	0.0

		EC TNFalpha + IL-1beta	
Primary Th1 rest	27.5	Bronchial epithelium TNFalpha + IL1beta	0.1
Primary Th2 rest	13.6	Small airway epithelium none	0.0
Primary Tr1 rest	32.8	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	3.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	8.5	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	5.8	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	3.1	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	2.9	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	4.7	KU-812 (Basophil) PMA/ionomycin	0.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	7.5	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	1.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	5.8	Liver cirrhosis	1.5
LAK cells IL-2+IL-12	2.3	Lupus kidney	0.6
LAK cells IL-2+IFN gamma	5.5	NCI-H292 none	2.5
LAK cells IL-2+ IL-18	5.5	NCI-H292 IL-4	1.8
LAK cells PMA/ionomycin	2.7	NCI-H292 IL-9	5.9
NK Cells IL-2 rest	6.0	NCI-H292 IL-13	2.3
Two Way MLR 3 day	2.1	NCI-H292 IFN gamma	3.0
Two Way MLR 5 day	0.9	HPAEC none	0.0
Two Way MLR 7 day	1.8	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	1.5	Lung fibroblast none	0.0
PBMC PWM	5.6	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.9	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.2	Lung fibroblast IL-13	0.0
B lymphocytes PWM	2.2	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L	3.7	Dermal fibroblast	0.0

620

and IL-4		CCD1070 rest	
EOL-1 dbcAMP	1.0	Dermal fibroblast CCD1070 TNF alpha	5.2
EOL-1 dbcAMP PMA/ionomycin	0.4	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.2	Dermal fibroblast IFN gamma	0.1
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.4
Monocytes rest	0.3	IBD Crohn's	8.4
Monocytes LPS	0.2	Colon	**100.0**
Macrophages rest	0.7	Lung	0.9
Macrophages LPS	0.0	Thymus	8.1
HUVEC none	0.0	Kidney	6.8
HUVEC starved	0.0		

Panel 1.3D Summary: Ag3372 Highest expression of the CG58553-01 gene is seen in the small intestine sample (CT=26.8). This gene encodes a novel vasopressin gene that plays a role in regulating electrolyte transport in the colon. Therefore, regulation of the transcript or the protein it encodes could be important in maintaining normal cellular homeostasis and in the treatment of Crohn's disease and ulcerative colitis.

Among tissues with metabolic function, this gene is expressed in liver and adipose. Thus, this gene product may be involved in disorders that affect these tissues, such as obesity and type II diabetes.

Low, but significant expression is also seen in the hippocampus. The hippocampus is critical for learning and memory. Thus, this gene product may have utility treating CNS disorders involving memory deficits, including Alzheimer's disease and aging.

References:

1. Sato Y, Hanai H, Nogaki A, Hirasawa K, Kaneko E, Hayashi H, Suzuki Y. Role of the vasopressin V(1) receptor in regulating the epithelial functions of the guinea pig distal colon. Am J Physiol 1999 Oct;277(4 Pt 1):G819-28.

Panel 4D Summary: Ag3372 In agreement with the results seen in panel 1.4, the highest level of expression of this gene is in the colon sample (CT=27.5). Interestingly, the

expression is significantly lower in the IBD colitis 2 (CT>35) and IBD Crohn's (CT=30.9)samples. Therefore, alterations in the expression of this gene may be used in the treatment of Crohn's disease and ulcerative colitis.

In addition, the expression of the CG58553-01 gene in several preparations of T lymphocytes suggests that small molecule antagonists, therapeutic antibodies specific for this molecule, or the extracellular domain of this protein, may be useful to reduce or eliminate the symptoms of Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

W. CG58626-01: Phospholipase

Expression of gene CG58626-01 was assessed using the primer-probe set Ag3386, described in Table WA. Results of the RTQ-PCR runs are shown in Tables WB, WC and WD.

Table WA. Probe Name Ag3386

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-agtggcggtcaaaacttactct-3'	22	1386	460
Probe	TET-5'-tggagacactgttgattccattactcctg-3'-TAMRA	29	1411	461
Reverse	5'-ctgctgttcagcatatccctta-3'	22	1455	462

Table WB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3386, Run 210154893	Tissue Name	Rel. Exp.(%) Ag3386, Run 210154893
AD 1 Hippo	6.4	Control (Path) 3 Temporal Ctx	4.2
AD 2 Hippo	21.5	Control (Path) 4 Temporal Ctx	25.0
AD 3 Hippo	5.0	AD 1 Occipital Ctx	14.4
AD 4 Hippo	5.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	95.3	AD 3 Occipital Ctx	3.3
AD 6 Hippo	51.1	AD 4 Occipital	18.7

		Ctx	
Control 2 Hippo	26.4	AD 5 Occipital Ctx	28.7
Control 4 Hippo	4.5	AD 6 Occipital Ctx	52.5
Control (Path) 3 Hippo	4.0	Control 1 Occipital Ctx	2.4
AD 1 Temporal Ctx	13.5	Control 2 Occipital Ctx	56.3
AD 2 Temporal Ctx	24.5	Control 3 Occipital Ctx	11.0
AD 3 Temporal Ctx	3.8	Control 4 Occipital Ctx	4.3
AD 4 Temporal Ctx	18.9	Control (Path) 1 Occipital Ctx	**100.0**
AD 5 Inf Temporal Ctx	95.9	Control (Path) 2 Occipital Ctx	8.8
AD 5 SupTemporal Ctx	37.6	Control (Path) 3 Occipital Ctx	1.7
AD 6 Inf Temporal Ctx	52.5	Control (Path) 4 Occipital Ctx	12.3
AD 6 Sup Temporal Ctx	63.7	Control 1 Parietal Ctx	5.4
Control 1 Temporal Ctx	4.9	Control 2 Parietal Ctx	39.5
Control 2 Temporal Ctx	38.4	Control 3 Parietal Ctx	11.3
Control 3 Temporal Ctx	12.2	Control (Path) 1 Parietal Ctx	77.4
Control 4 Temporal Ctx	5.0	Control (Path) 2 Parietal Ctx	20.7
Control (Path) 1 Temporal Ctx	76.8	Control (Path) 3 Parietal Ctx	2.7
Control (Path) 2 Temporal Ctx	37.6	Control (Path) 4 Parietal Ctx	45.4

Table WC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3386, Run 217043839	Tissue Name	Rel. Exp.(%) Ag3386, Run 217043839
Adipose	12.2	Renal ca. TK-10	10.7
Melanoma* Hs688(A).T	26.4	Bladder	18.4
Melanoma*	30.4	Gastric ca. (liver met.)	26.6

Hs688(B).T		NCI-N87	
Melanoma* M14	13.1	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	22.8	Colon ca. SW-948	10.8
Melanoma* SK-MEL-5	22.7	Colon ca. SW480	40.9
Squamous cell carcinoma SCC-4	11.2	Colon ca.* (SW480 met) SW620	20.4
Testis Pool	47.0	Colon ca. HT29	5.2
Prostate ca.* (bone met) PC-3	80.1	Colon ca. HCT-116	**100.0**
Prostate Pool	7.1	Colon ca. CaCo-2	13.8
Placenta	3.2	Colon cancer tissue	13.6
Uterus Pool	6.4	Colon ca. SW1116	10.2
Ovarian ca. OVCAR-3	22.8	Colon ca. Colo-205	1.8
Ovarian ca. SK-OV-3	94.0	Colon ca. SW-48	2.4
Ovarian ca. OVCAR-4	4.7	Colon Pool	27.7
Ovarian ca. OVCAR-5	29.3	Small Intestine Pool	14.6
Ovarian ca. IGROV-1	12.7	Stomach Pool	12.2
Ovarian ca. OVCAR-8	11.1	Bone Marrow Pool	6.9
Ovary	11.6	Fetal Heart	8.1
Breast ca. MCF-7	36.9	Heart Pool	6.3
Breast ca. MDA-MB-231	39.5	Lymph Node Pool	13.9
Breast ca. BT 549	28.5	Fetal Skeletal Muscle	3.6
Breast ca. T47D	52.9	Skeletal Muscle Pool	6.7
Breast ca. MDA-N	11.3	Spleen Pool	17.1
Breast Pool	28.1	Thymus Pool	26.1
Trachea	11.0	CNS cancer (glio/astro) U87-MG	33.2
Lung	6.0	CNS cancer (glio/astro) U-118-MG	44.1
Fetal Lung	39.2	CNS cancer (neuro;met) SK-N-AS	44.4
Lung ca. NCI-N417	6.3	CNS cancer (astro) SF-539	10.4
Lung ca. LX-1	33.9	CNS cancer (astro) SNB-75	27.7

Lung ca. NCI-H146	14.3	CNS cancer (glio) SNB-19	10.2
Lung ca. SHP-77	73.2	CNS cancer (glio) SF-295	28.7
Lung ca. A549	25.3	Brain (Amygdala) Pool	23.2
Lung ca. NCI-H526	5.8	Brain (cerebellum)	19.8
Lung ca. NCI-H23	30.1	Brain (fetal)	35.6
Lung ca. NCI-H460	20.2	Brain (Hippocampus) Pool	25.2
Lung ca. HOP-62	11.9	Cerebral Cortex Pool	39.2
Lung ca. NCI-H522	20.7	Brain (Substantia nigra) Pool	23.0
Liver	0.7	Brain (Thalamus) Pool	45.7
Fetal Liver	29.5	Brain (whole)	24.0
Liver ca. HepG2	10.1	Spinal Cord Pool	22.5
Kidney Pool	21.3	Adrenal Gland	8.5
Fetal Kidney	19.5	Pituitary gland Pool	7.0
Renal ca. 786-0	15.9	Salivary Gland	1.9
Renal ca. A498	3.5	Thyroid (female)	3.2
Renal ca. ACHN	8.0	Pancreatic ca. CAPAN2	3.7
Renal ca. UO-31	12.2	Pancreas Pool	18.2

Table WD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3386, Run 165296474	Tissue Name	Rel. Exp.(%) Ag3386, Run 165296474
Secondary Th1 act	30.4	HUVEC IL-1beta	2.0
Secondary Th2 act	35.6	HUVEC IFN gamma	3.3
Secondary Tr1 act	27.9	HUVEC TNF alpha + IFN gamma	3.8
Secondary Th1 rest	8.9	HUVEC TNF alpha + IL4	3.3
Secondary Th2 rest	8.0	HUVEC IL-11	1.5
Secondary Tr1 rest	11.3	Lung Microvascular EC none	5.5
Primary Th1 act	57.4	Lung Microvascular EC TNFalpha + IL-1beta	4.8
Primary Th2 act	36.9	Microvascular Dermal EC none	3.7
Primary Tr1 act	51.1	Microsvasular Dermal EC TNFalpha + IL-1beta	3.3
Primary Th1 rest	54.0	Bronchial epithelium	5.5

625

		TNFalpha + IL1beta	
Primary Th2 rest	18.8	Small airway epithelium none	2.1
Primary Tr1 rest	24.7	Small airway epithelium TNFalpha + IL-1beta	6.1
CD45RA CD4 lymphocyte act	12.4	Coronery artery SMC rest	4.6
CD45RO CD4 lymphocyte act	33.9	Coronery artery SMC TNFalpha + IL-1beta	2.4
CD8 lymphocyte act	29.3	Astrocytes rest	3.0
Secondary CD8 lymphocyte rest	26.1	Astrocytes TNFalpha + IL-1beta	3.0
Secondary CD8 lymphocyte act	20.7	KU-812 (Basophil) rest	12.1
CD4 lymphocyte none	1.1	KU-812 (Basophil) PMA/ionomycin	27.7
2ry Th1/Th2/Tr1_anti-CD95 CH11	12.8	CCD1106 (Keratinocytes) none	2.7
LAK cells rest	12.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.8
LAK cells IL-2	24.3	Liver cirrhosis	0.3
LAK cells IL-2+IL-12	28.7	Lupus kidney	0.7
LAK cells IL-2+IFN gamma	42.0	NCI-H292 none	8.1
LAK cells IL-2+ IL-18	45.1	NCI-H292 IL-4	9.5
LAK cells PMA/ionomycin	8.8	NCI-H292 IL-9	8.5
NK Cells IL-2 rest	21.8	NCI-H292 IL-13	4.5
Two Way MLR 3 day	18.7	NCI-H292 IFN gamma	3.6
Two Way MLR 5 day	11.0	HPAEC none	2.5
Two Way MLR 7 day	10.9	HPAEC TNF alpha + IL-1 beta	3.0
PBMC rest	4.5	Lung fibroblast none	4.6
PBMC PWM	66.0	Lung fibroblast TNF alpha + IL-1 beta	3.3
PBMC PHA-L	17.9	Lung fibroblast IL-4	12.1
Ramos (B cell) none	26.1	Lung fibroblast IL-9	12.3
Ramos (B cell) ionomycin	**100.0**	Lung fibroblast IL-13	6.7
B lymphocytes PWM	88.9	Lung fibroblast IFN gamma	16.6
B lymphocytes CD40L and IL-4	49.3	Dermal fibroblast CCD1070 rest	8.2
EOL-1 dbcAMP	13.0	Dermal fibroblast	37.1

		CCD1070 TNF alpha	
EOL-1 dbcAMP PMA/ionomycin	9.5	Dermal fibroblast CCD1070 IL-1 beta	4.4
Dendritic cells none	8.9	Dermal fibroblast IFN gamma	6.0
Dendritic cells LPS	5.4	Dermal fibroblast IL-4	12.1
Dendritic cells anti-CD40	4.3	IBD Colitis 2	0.7
Monocytes rest	4.7	IBD Crohn's	0.5
Monocytes LPS	2.6	Colon	3.4
Macrophages rest	8.8	Lung	4.9
Macrophages LPS	2.8	Thymus	4.1
HUVEC none	2.8	Kidney	13.0
HUVEC starved	6.4		

CNS_neurodegeneration_v1.0 Summary: Ag3386 This panel confirms the expression of this gene at moderate to low levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.4 Summary: Ag3386 This gene is moderately expressed in most of the samples on this panel. Based on expression in this panel, this gene may be involved in gastric, pancreatic, brain, colon, renal, lung, breast, ovarian and prostate cancer as well as melanomas. Thus, expression of this gene could be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition using antibodies or small molecule drugs might be of use in the treatment of these cancers.

This gene product is also expressed in adipose, pancreas, adrenal, thyroid, pituitary, skeletal muscle, heart, and liver. This widespread expression in tissues with metabolic function suggests that this gene product may be important for the pathogenesis, diagnosis, and/or treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

In addition, this gene is expressed at moderate levels in the CNS. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

627

Panel 4D Summary: Ag3386 The CG58626-01 transcript is expressed ubiquitously in this panel. Highest expression of this transcript is seen in activated Ramos cells and activated B cells (CTs=27). The expression of this transcript in activated lymphoid cells when compared to non activated cells suggests that the CG58626-01 gene may be important for the diagnosis or pathogenesis of immune mediated diseases. Therefore, modulation of the expression and/or activity of this gene product might important for the treatment of autoimmune diseases, allergy, and delayed type hypersensitivity.

X. CG57597-01: Hypothetical protein

Expression of gene CG57597-01 was assessed using the primer-probe set Ag3293, described in Table XA.

Table XA. Probe Name Ag3293

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cagaaacctgtgaactctgcat-3'	22	40	463
Probe	TET-5'-atgcaccaccactcctggctaatttt-3'-TAMRA	26	69	464
Reverse	5'-ataaaaggtttgagccggatt-3'	21	115	465

CNS_neurodegeneration_v1.0 Summary: Ag3293 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3293 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag3293 - Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Y. CG57804-01: talin

Expression of gene CG57804-01 was assessed using the primer-probe set Ag3337, described in Table YA. Results of the RTQ-PCR runs are shown in Tables YB, YC and YD.

Table YA. Probe Name Ag3337

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggatttcaagcccagatacaat-3'	22	781	466
Probe	TET-5'-tggacctcatgtggaacataaacaca-3'-TAMRA	26	804	467
Reverse	5'-ggcaggaattccttcagatc-3'	20	844	468

Table YB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3337, Run 210138775	Tissue Name	Rel. Exp.(%) Ag3337, Run 210138775
AD 1 Hippo	6.8	Control (Path) 3 Temporal Ctx	3.6
AD 2 Hippo	25.3	Control (Path) 4 Temporal Ctx	22.4
AD 3 Hippo	3.6	AD 1 Occipital Ctx	5.6
AD 4 Hippo	5.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	78.5	AD 3 Occipital Ctx	2.2
AD 6 Hippo	27.5	AD 4 Occipital Ctx	14.7
Control 2 Hippo	27.4	AD 5 Occipital Ctx	44.1
Control 4 Hippo	8.1	AD 6 Occipital Ctx	16.6
Control (Path) 3 Hippo	4.3	Control 1 Occipital Ctx	1.6
AD 1 Temporal Ctx	7.6	Control 2 Occipital Ctx	67.8
AD 2 Temporal Ctx	24.5	Control 3 Occipital Ctx	11.9
AD 3 Temporal Ctx	3.3	Control 4 Occipital Ctx	3.0
AD 4 Temporal Ctx	15.3	Control (Path) 1 Occipital Ctx	89.5
AD 5 Inf Temporal Ctx	89.5	Control (Path) 2 Occipital Ctx	8.2
AD 5 Sup Temporal Ctx	35.8	Control (Path) 3 Occipital Ctx	0.6
AD 6 Inf Temporal Ctx	27.4	Control (Path) 4 Occipital Ctx	10.3
AD 6 Sup Temporal Ctx	32.8	Control 1 Parietal Ctx	3.6
Control 1 Temporal Ctx	3.4	Control 2 Parietal Ctx	23.7
Control 2 Temporal Ctx	47.6	Control 3 Parietal Ctx	14.1
Control 3	12.4	Control (Path) 1	100.0

629

Temporal Ctx		Parietal Ctx	
Control 3 Temporal Ctx	5.8	Control (Path) 2 Parietal Ctx	21.9
Control (Path) 1 Temporal Ctx	64.2	Control (Path) 3 Parietal Ctx	2.0
Control (Path) 2 Temporal Ctx	42.0	Control (Path) 4 Parietal Ctx	39.2

Table YC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3337, Run 215773748	Tissue Name	Rel. Exp.(%) Ag3337, Run 215773748
Adipose	20.2	Renal ca. TK-10	22.1
Melanoma* Hs688(A).T	58.6	Bladder	14.2
Melanoma* Hs688(B).T	22.8	Gastric ca. (liver met.) NCI-N87	16.2
Melanoma* M14	5.7	Gastric ca. KATO III	**100.0**
Melanoma* LOXIMVI	5.5	Colon ca. SW-948	16.3
Melanoma* SK-MEL-5	3.4	Colon ca. SW480	4.2
Squamous cell carcinoma SCC-4	4.4	Colon ca.* (SW480 met) SW620	2.6
Testis Pool	5.1	Colon ca. HT29	0.7
Prostate ca.* (bone met) PC-3	6.4	Colon ca. HCT-116	7.6
Prostate Pool	3.4	Colon ca. CaCo-2	81.8
Placenta	1.6	Colon cancer tissue	1.7
Uterus Pool	2.1	Colon ca. SW1116	1.6
Ovarian ca. OVCAR-3	8.9	Colon ca. Colo-205	0.1
Ovarian ca. SK-OV-3	32.1	Colon ca. SW-48	3.2
Ovarian ca. OVCAR-4	7.2	Colon Pool	8.0
Ovarian ca. OVCAR-5	21.0	Small Intestine Pool	7.9
Ovarian ca. IGROV-1	23.5	Stomach Pool	5.7
Ovarian ca. OVCAR-8	5.4	Bone Marrow Pool	3.8
Ovary	11.7	Fetal Heart	24.8
Breast ca. MCF-7	5.1	Heart Pool	10.2

Breast ca. MDA-MB-231	19.5	Lymph Node Pool	10.6
Breast ca. BT 549	11.7	Fetal Skeletal Muscle	30.1
Breast ca. T47D	30.8	Skeletal Muscle Pool	24.8
Breast ca. MDA-N	5.0	Spleen Pool	2.9
Breast Pool	6.9	Thymus Pool	0.0
Trachea	10.3	CNS cancer (glio/astro) U87-MG	10.7
Lung	2.2	CNS cancer (glio/astro) U-118-MG	68.3
Fetal Lung	10.8	CNS cancer (neuro;met) SK-N-AS	23.5
Lung ca. NCI-N417	0.8	CNS cancer (astro) SF-539	21.5
Lung ca. LX-1	1.7	CNS cancer (astro) SNB-75	40.3
Lung ca. NCI-H146	0.4	CNS cancer (glio) SNB-19	27.7
Lung ca. SHP-77	11.9	CNS cancer (glio) SF-295	38.2
Lung ca. A549	13.6	Brain (Amygdala) Pool	28.7
Lung ca. NCI-H526	7.6	Brain (cerebellum)	38.7
Lung ca. NCI-H23	10.2	Brain (fetal)	58.6
Lung ca. NCI-H460	5.1	Brain (Hippocampus) Pool	25.7
Lung ca. HOP-62	3.8	Cerebral Cortex Pool	59.0
Lung ca. NCI-H522	10.1	Brain (Substantia nigra) Pool	39.2
Liver	0.3	Brain (Thalamus) Pool	51.4
Fetal Liver	15.3	Brain (whole)	58.2
Liver ca. HepG2	53.2	Spinal Cord Pool	18.6
Kidney Pool	18.4	Adrenal Gland	11.1
Fetal Kidney	11.4	Pituitary gland Pool	4.7
Renal ca. 786-0	31.0	Salivary Gland	14.0
Renal ca. A498	0.7	Thyroid (female)	4.9
Renal ca. ACHN	20.3	Pancreatic ca. CAPAN2	7.5
Renal ca. UO-31	8.1	Pancreas Pool	9.4

Table YD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3337, Run 165725932	Tissue Name	Rel. Exp.(%) Ag3337, Run 165725932

Secondary Th1 act	0.0	HUVEC IL-1beta	0.7
Secondary Th2 act	0.0	HUVEC IFN gamma	3.9
Secondary Tr1 act	0.4	HUVEC TNF alpha + IFN gamma	0.3
Secondary Th1 rest	0.4	HUVEC TNF alpha + IL4	0.6
Secondary Th2 rest	0.0	HUVEC IL-11	0.3
Secondary Tr1 rest	0.3	Lung Microvascular EC none	2.1
Primary Th1 act	0.3	Lung Microvascular EC TNFalpha + IL-1beta	5.1
Primary Th2 act	1.3	Microvascular Dermal EC none	16.4
Primary Tr1 act	0.6	Microsvasular Dermal EC TNFalpha + IL-1beta	9.8
Primary Th1 rest	1.3	Bronchial epithelium TNFalpha + IL1beta	1.2
Primary Th2 rest	0.6	Small airway epithelium none	1.3
Primary Tr1 rest	0.3	Small airway epithelium TNFalpha + IL-1beta	2.1
CD45RA CD4 lymphocyte act	18.7	Coronery artery SMC rest	9.9
CD45RO CD4 lymphocyte act	0.6	Coronery artery SMC TNFalpha + IL-1beta	3.5
CD8 lymphocyte act	1.2	Astrocytes rest	**100.0**
Secondary CD8 lymphocyte rest	0.9	Astrocytes TNFalpha + IL-1beta	65.5
Secondary CD8 lymphocyte act	0.2	KU-812 (Basophil) rest	11.7
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	8.5
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.3	CCD1106 (Keratinocytes) none	2.0
LAK cells rest	4.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	2.0
LAK cells IL-2	1.2	Liver cirrhosis	3.6
LAK cells IL-2+IL-12	0.4	Lupus kidney	13.6
LAK cells IL-2+IFN gamma	2.1	NCI-H292 none	11.0
LAK cells IL-2+ IL-18	1.2	NCI-H292 IL-4	25.0
LAK cells PMA/ionomycin	2.0	NCI-H292 IL-9	15.6
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	12.5

Two Way MLR 3 day	5.2	NCI-H292 IFN gamma	4.6
Two Way MLR 5 day	2.7	HPAEC none	1.5
Two Way MLR 7 day	1.8	HPAEC TNF alpha + IL-1 beta	2.5
PBMC rest	0.2	Lung fibroblast none	80.1
PBMC PWM	1.9	Lung fibroblast TNF alpha + IL-1 beta	22.7
PBMC PHA-L	0.3	Lung fibroblast IL-4	97.3
Ramos (B cell) none	2.4	Lung fibroblast IL-9	47.6
Ramos (B cell) ionomycin	2.1	Lung fibroblast IL-13	81.8
B lymphocytes PWM	0.7	Lung fibroblast IFN gamma	50.7
B lymphocytes CD40L and IL-4	0.6	Dermal fibroblast CCD1070 rest	94.6
EOL-1 dbcAMP	4.9	Dermal fibroblast CCD1070 TNF alpha	43.2
EOL-1 dbcAMP PMA/ionomycin	1.2	Dermal fibroblast CCD1070 IL-1 beta	31.2
Dendritic cells none	12.8	Dermal fibroblast IFN gamma	14.2
Dendritic cells LPS	1.3	Dermal fibroblast IL-4	95.9
Dendritic cells anti-CD40	11.4	IBD Colitis 2	1.2
Monocytes rest	0.5	IBD Crohn's	9.1
Monocytes LPS	1.3	Colon	60.7
Macrophages rest	13.6	Lung	8.0
Macrophages LPS	2.8	Thymus	39.2
HUVEC none	1.0	Kidney	10.4
HUVEC starved	1.6		

CNS_neurodegeneration_v1.0 Summary: Ag3337 - This panel confirms the expression of this gene at low to moderate levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders

General_screening_panel_v1.4 Summary: Ag3337 - This gene is expressed in almost all samples on this panel. This gene is expressed at moderate levels in the CNS. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

In addition, this gene is also expressed in adipose, pancreas, adrenal, thyroid, pituitary, skeletal muscle, heart, and liver. This widespread expression in tissues with metabolic function suggests that this gene product may be important for the pathogenesis, diagnosis, and/or treatment of metabolic and endocrine diseases, including obesity and

5 Types 1 and 2 diabetes.

Panel 4D Summary: Ag3337 This gene is most highly expressed in resting astrocytes (CT=28.9). In addition, this gene is highly expressed in a cluster of treated and untreated samples derived from lung and dermal fibroblasts. Thus, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pathological and

10 inflammatory lung and skin diseases, such as psoriasis, asthma, emphysema, and allergies.

Z. CG57551-01: NAC-1 Like Gene

Expression of gene CG57551-01 was assessed using the primer-probe set Ag3282, described in Table ZA. Results of the RTQ-PCR runs are shown in Tables ZB, ZC and ZD.

Table ZA. Probe Name Ag3282

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cagatcctcagcttctgctaca-3'	22	269	469
Probe	TET-5'-accagttcctgctcatgtacacggct-3'-TAMRA	26	318	470
Reverse	5'-atctcctggatctgcaggaa-3'	20	347	471

15 Table ZB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3282, Run 210060482	Tissue Name	Rel. Exp.(%) Ag3282, Run 210060482
AD 1 Hippo	22.8	Control (Path) 3 Temporal Ctx	9.7
AD 2 Hippo	49.0	Control (Path) 4 Temporal Ctx	24.3
AD 3 Hippo	11.5	AD 1 Occipital Ctx	16.5
AD 4 Hippo	12.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	66.9	AD 3 Occipital Ctx	10.2
AD 6 Hippo	59.9	AD 4 Occipital	18.2

634

		Ctx	
Control 2 Hippo	49.3	AD 5 Occipital Ctx	9.5
Control 4 Hippo	18.7	AD 6 Occipital Ctx	41.5
Control (Path) 3 Hippo	6.3	Control 1 Occipital Ctx	6.8
AD 1 Temporal Ctx	19.2	Control 2 Occipital Ctx	91.4
AD 2 Temporal Ctx	40.3	Control 3 Occipital Ctx	16.3
AD 3 Temporal Ctx	14.3	Control 4 Occipital Ctx	12.2
AD 4 Temporal Ctx	18.3	Control (Path) 1 Occipital Ctx	**100.0**
AD 5 Inf Temporal Ctx	66.0	Control (Path) 2 Occipital Ctx	9.2
AD 5 SupTemporal Ctx	37.4	Control (Path) 3 Occipital Ctx	5.3
AD 6 Inf Temporal Ctx	36.1	Control (Path) 4 Occipital Ctx	15.8
AD 6 Sup Temporal Ctx	34.4	Control 1 Parietal Ctx	11.7
Control 1 Temporal Ctx	10.0	Control 2 Parietal Ctx	34.2
Control 2 Temporal Ctx	74.7	Control 3 Parietal Ctx	23.3
Control 3 Temporal Ctx	15.0	Control (Path) 1 Parietal Ctx	72.7
Control 4 Temporal Ctx	15.5	Control (Path) 2 Parietal Ctx	21.6
Control (Path) 1 Temporal Ctx	74.2	Control (Path) 3 Parietal Ctx	5.5
Control (Path) 2 Temporal Ctx	31.2	Control (Path) 4 Parietal Ctx	35.8

Table ZC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3282, Run 216512995	Tissue Name	Rel. Exp.(%) Ag3282, Run 216512995
Adipose	1.8	Renal ca. TK-10	22.7
Melanoma* Hs688(A).T	16.3	Bladder	6.3
Melanoma*	25.0	Gastric ca. (liver met.)	47.0

Hs688(B).T		NCI-N87	
Melanoma* M14	25.3	Gastric ca. KATO III	45.7
Melanoma* LOXIMVI	21.6	Colon ca. SW-948	19.3
Melanoma* SK-MEL-5	17.0	Colon ca. SW480	50.3
Squamous cell carcinoma SCC-4	24.7	Colon ca.* (SW480 met) SW620	25.9
Testis Pool	6.1	Colon ca. HT29	17.7
Prostate ca.* (bone met) PC-3	67.8	Colon ca. HCT-116	**100.0**
Prostate Pool	3.5	Colon ca. CaCo-2	29.1
Placenta	9.6	Colon cancer tissue	14.0
Uterus Pool	0.6	Colon ca. SW1116	12.7
Ovarian ca. OVCAR-3	41.2	Colon ca. Colo-205	7.6
Ovarian ca. SK-OV-3	65.5	Colon ca. SW-48	5.8
Ovarian ca. OVCAR-4	35.8	Colon Pool	4.9
Ovarian ca. OVCAR-5	37.6	Small Intestine Pool	2.4
Ovarian ca. IGROV-1	28.9	Stomach Pool	3.3
Ovarian ca. OVCAR-8	14.2	Bone Marrow Pool	1.5
Ovary	3.9	Fetal Heart	3.0
Breast ca. MCF-7	42.3	Heart Pool	2.2
Breast ca. MDA-MB-231	69.7	Lymph Node Pool	5.6
Breast ca. BT 549	51.4	Fetal Skeletal Muscle	1.5
Breast ca. T47D	86.5	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	26.4	Spleen Pool	2.8
Breast Pool	4.6	Thymus Pool	3.8
Trachea	8.7	CNS cancer (glio/astro) U87-MG	60.3
Lung	0.2	CNS cancer (glio/astro) U-118-MG	**100.0**
Fetal Lung	6.3	CNS cancer (neuro;met) SK-N-AS	47.3
Lung ca. NCI-N417	8.4	CNS cancer (astro) SF-539	22.8
Lung ca. LX-1	17.3	CNS cancer (astro) SNB-75	47.3

636

Lung ca. NCI-H146	15.3	CNS cancer (glio) SNB-19	29.3
Lung ca. SHP-77	16.5	CNS cancer (glio) SF-295	49.3
Lung ca. A549	27.2	Brain (Amygdala) Pool	6.9
Lung ca. NCI-H526	6.1	Brain (cerebellum)	15.1
Lung ca. NCI-H23	25.9	Brain (fetal)	9.2
Lung ca. NCI-H460	8.0	Brain (Hippocampus) Pool	8.9
Lung ca. HOP-62	11.9	Cerebral Cortex Pool	13.4
Lung ca. NCI-H522	21.9	Brain (Substantia nigra) Pool	15.3
Liver	1.7	Brain (Thalamus) Pool	11.5
Fetal Liver	9.8	Brain (whole)	12.6
Liver ca. HepG2	18.2	Spinal Cord Pool	7.1
Kidney Pool	5.0	Adrenal Gland	5.3
Fetal Kidney	3.4	Pituitary gland Pool	1.9
Renal ca. 786-0	42.0	Salivary Gland	4.0
Renal ca. A498	16.7	Thyroid (female)	3.6
Renal ca. ACHN	13.9	Pancreatic ca. CAPAN2	15.5
Renal ca. UO-31	17.4	Pancreas Pool	6.6

Table ZD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3282, Run 164634321	Tissue Name	Rel. Exp.(%) Ag3282, Run 164634321
Secondary Th1 act	52.9	HUVEC IL-1beta	13.6
Secondary Th2 act	67.8	HUVEC IFN gamma	42.9
Secondary Tr1 act	75.3	HUVEC TNF alpha + IFN gamma	37.1
Secondary Th1 rest	8.4	HUVEC TNF alpha + IL4	42.6
Secondary Th2 rest	11.4	HUVEC IL-11	25.9
Secondary Tr1 rest	12.2	Lung Microvascular EC none	41.2
Primary Th1 act	53.6	Lung Microvascular EC TNFalpha + IL-1beta	36.3
Primary Th2 act	44.4	Microvascular Dermal EC none	50.3
Primary Tr1 act	60.7	Microsvasular Dermal EC TNFalpha + IL-1beta	33.0
Primary Th1 rest	37.6	Bronchial epithelium	51.4

		TNFalpha + IL1beta	
Primary Th2 rest	15.8	Small airway epithelium none	23.3
Primary Tr1 rest	18.3	Small airway epithelium TNFalpha + IL-1beta	71.7
CD45RA CD4 lymphocyte act	33.0	Coronery artery SMC rest	43.5
CD45RO CD4 lymphocyte act	54.7	Coronery artery SMC TNFalpha + IL-1beta	31.0
CD8 lymphocyte act	42.9	Astrocytes rest	38.4
Secondary CD8 lymphocyte rest	50.3	Astrocytes TNFalpha + IL-1beta	37.1
Secondary CD8 lymphocyte act	32.5	KU-812 (Basophil) rest	36.1
CD4 lymphocyte none	2.4	KU-812 (Basophil) PMA/ionomycin	90.8
2ry Th1/Th2/Tr1_anti-CD95 CH11	11.7	CCD1106 (Keratinocytes) none	64.2
LAK cells rest	18.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	34.4
LAK cells IL-2	41.2	Liver cirrhosis	2.9
LAK cells IL-2+IL-12	29.5	Lupus kidney	2.2
LAK cells IL-2+IFN gamma	36.3	NCI-H292 none	38.4
LAK cells IL-2+ IL-18	34.2	NCI-H292 IL-4	66.9
LAK cells PMA/ionomycin	11.8	NCI-H292 IL-9	62.4
NK Cells IL-2 rest	29.3	NCI-H292 IL-13	65.1
Two Way MLR 3 day	21.9	NCI-H292 IFN gamma	48.3
Two Way MLR 5 day	27.7	HPAEC none	31.2
Two Way MLR 7 day	27.0	HPAEC TNF alpha + IL-1 beta	37.6
PBMC rest	6.5	Lung fibroblast none	35.6
PBMC PWM	89.5	Lung fibroblast TNF alpha + IL-1 beta	20.7
PBMC PHA-L	53.6	Lung fibroblast IL-4	63.3
Ramos (B cell) none	40.6	Lung fibroblast IL-9	55.5
Ramos (B cell) ionomycin	56.3	Lung fibroblast IL-13	44.8
B lymphocytes PWM	**100.0**	Lung fibroblast IFN gamma	71.2
B lymphocytes CD40L and IL-4	41.2	Dermal fibroblast CCD1070 rest	78.5
EOL-1 dbcAMP	50.0	Dermal fibroblast	88.9

638

		CCD1070 TNF alpha	
EOL-1 dbcAMP PMA/ionomycin	46.3	Dermal fibroblast CCD1070 IL-1 beta	49.7
Dendritic cells none	33.2	Dermal fibroblast IFN gamma	21.5
Dendritic cells LPS	26.1	Dermal fibroblast IL-4	43.8
Dendritic cells anti-CD40	29.9	IBD Colitis 2	1.2
Monocytes rest	17.1	IBD Crohn's	1.8
Monocytes LPS	14.0	Colon	15.4
Macrophages rest	59.0	Lung	16.6
Macrophages LPS	29.1	Thymus	15.6
HUVEC none	35.1	Kidney	18.9
HUVEC starved	62.0		

CNS_neurodegeneration_v1.0 Summary: Ag3282 - This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.4 Summary: Ag3282 Highest expression of this gene is seen in a brain cancer cell line (CT=24.3). This gene appears to be expressed more highly in the cancer cell lines than in the normal tissue samples on this panel and may be involved in cellular growth and proliferation. Based on this expression profile, this gene may be involved in gastric, pancreatic, brain, colon, renal, lung, breast, ovarian and prostate cancer as well as melanomas. Thus, expression of this gene could be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition using antibodies or small molecule drugs might be of use in the treatment of these cancers.

This gene is also expressed at high levels in all regions of the CNS examined. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

In addition, this gene product is expressed in adipose, pancreas, adrenal, thyroid, pituitary, fetal skeletal muscle, heart, and liver. This widespread expression in tissues with

metabolic function suggests that this gene product may be important for the pathogenesis, diagnosis, and/or treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

Furthermore, this gene is more highly expressed in fetal skeletal muscle (CT=30.4) and liver (CT=27) when compared to expression in the adult skeletal muscle (CT>35) and liver (CT=30) may be useful for the differentiation of the fetal and adult sources of this tissue.

Panel 4D Summary: Ag3282 This gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease. Highest expression is seen in polkweed mitogen stimulated B lymphocytes (CT=25.7). In addition, expression is seen in members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in Panel 1.4 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

AA. CG57411-01: KELCH-LIKE PROTEIN KLHL3C

Expression of gene CG57411-01 was assessed using the primer-probe set Ag3229, described in Table AAA. Results of the RTQ-PCR runs are shown in Tables AAB, AAC, AAD and AAE.

Table AAA. Probe Name Ag3229

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gcagcgagctctaccacat-3'	19	287	472

Probe	TET-5'-aaggccttcgcgctgcagatctt-3'-TAMRA	23	310	473
Reverse	5'-aagtcgtccttggagatgct-3'	20	364	474

Table AAB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3229, Run 209862301	Tissue Name	Rel. Exp.(%) Ag3229, Run 209862301
AD 1 Hippo	16.3	Control (Path) 3 Temporal Ctx	8.0
AD 2 Hippo	34.6	Control (Path) 4 Temporal Ctx	35.8
AD 3 Hippo	15.9	AD 1 Occipital Ctx	18.6
AD 4 Hippo	6.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	**100.0**	AD 3 Occipital Ctx	11.7
AD 6 Hippo	35.4	AD 4 Occipital Ctx	17.7
Control 2 Hippo	31.2	AD 5 Occipital Ctx	49.7
Control 4 Hippo	12.1	AD 6 Occipital Ctx	14.2
Control (Path) 3 Hippo	6.2	Control 1 Occipital Ctx	3.3
AD 1 Temporal Ctx	21.6	Control 2 Occipital Ctx	69.3
AD 2 Temporal Ctx	33.0	Control 3 Occipital Ctx	26.6
AD 3 Temporal Ctx	14.1	Control 4 Occipital Ctx	7.5
AD 4 Temporal Ctx	16.8	Control (Path) 1 Occipital Ctx	72.2
AD 5 Inf Temporal Ctx	71.7	Control (Path) 2 Occipital Ctx	13.7
AD 5 Sup Temporal Ctx	32.3	Control (Path) 3 Occipital Ctx	6.3
AD 6 Inf Temporal Ctx	30.6	Control (Path) 4 Occipital Ctx	16.8
AD 6 Sup Temporal Ctx	33.9	Control 1 Parietal Ctx	8.6
Control 1 Temporal Ctx	4.4	Control 2 Parietal Ctx	39.8
Control 2 Temporal Ctx	56.6	Control 3 Parietal Ctx	21.5
Control 3 Temporal Ctx	19.6	Control (Path) 1 Parietal Ctx	66.4
Control 3 Temporal Ctx	14.2	Control (Path) 2 Parietal Ctx	26.8

641

Control (Path) 1 Temporal Ctx	62.0	Control (Path) 3 Parietal Ctx	5.2
Control (Path) 2 Temporal Ctx	36.1	Control (Path) 4 Parietal Ctx	54.3

Table AAC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3229, Run 214439727	Tissue Name	Rel. Exp.(%) Ag3229, Run 214439727
Adipose	6.0	Renal ca. TK-10	20.4
Melanoma* Hs688(A).T	8.1	Bladder	6.7
Melanoma* Hs688(B).T	13.5	Gastric ca. (liver met.) NCI-N87	11.2
Melanoma* M14	2.1	Gastric ca. KATO III	59.5
Melanoma* LOXIMVI	24.8	Colon ca. SW-948	0.6
Melanoma* SK-MEL-5	20.7	Colon ca. SW480	31.6
Squamous cell carcinoma SCC-4	6.7	Colon ca.* (SW480 met) SW620	4.7
Testis Pool	3.0	Colon ca. HT29	2.7
Prostate ca.* (bone met) PC-3	17.8	Colon ca. HCT-116	35.1
Prostate Pool	8.5	Colon ca. CaCo-2	2.5
Placenta	14.2	Colon cancer tissue	6.2
Uterus Pool	5.8	Colon ca. SW1116	3.3
Ovarian ca. OVCAR-3	40.9	Colon ca. Colo-205	5.4
Ovarian ca. SK-OV-3	17.7	Colon ca. SW-48	3.3
Ovarian ca. OVCAR-4	11.9	Colon Pool	25.0
Ovarian ca. OVCAR-5	84.1	Small Intestine Pool	14.9
Ovarian ca. IGROV-1	2.0	Stomach Pool	6.4
Ovarian ca. OVCAR-8	8.1	Bone Marrow Pool	9.7
Ovary	8.7	Fetal Heart	1.7
Breast ca. MCF-7	0.9	Heart Pool	10.7
Breast ca. MDA-MB-231	30.1	Lymph Node Pool	21.8
Breast ca. BT 549	8.1	Fetal Skeletal Muscle	4.2

Breast ca. T47D	**100.0**	Skeletal Muscle Pool	8.7
Breast ca. MDA-N	0.0	Spleen Pool	10.4
Breast Pool	22.4	Thymus Pool	11.2
Trachea	10.4	CNS cancer (glio/astro) U87-MG	55.5
Lung	1.7	CNS cancer (glio/astro) U-118-MG	44.8
Fetal Lung	6.7	CNS cancer (neuro;met) SK-N-AS	5.8
Lung ca. NCI-N417	11.7	CNS cancer (astro) SF-539	0.4
Lung ca. LX-1	37.1	CNS cancer (astro) SNB-75	5.0
Lung ca. NCI-H146	6.2	CNS cancer (glio) SNB-19	2.9
Lung ca. SHP-77	61.1	CNS cancer (glio) SF-295	39.0
Lung ca. A549	6.3	Brain (Amygdala) Pool	8.4
Lung ca. NCI-H526	8.7	Brain (cerebellum)	22.2
Lung ca. NCI-H23	6.3	Brain (fetal)	48.6
Lung ca. NCI-H460	2.9	Brain (Hippocampus) Pool	8.5
Lung ca. HOP-62	8.1	Cerebral Cortex Pool	20.7
Lung ca. NCI-H522	0.5	Brain (Substantia nigra) Pool	14.7
Liver	0.3	Brain (Thalamus) Pool	13.8
Fetal Liver	1.8	Brain (whole)	11.5
Liver ca. HepG2	0.2	Spinal Cord Pool	3.3
Kidney Pool	26.8	Adrenal Gland	29.9
Fetal Kidney	10.2	Pituitary gland Pool	10.7
Renal ca. 786-0	7.5	Salivary Gland	4.1
Renal ca. A498	4.0	Thyroid (female)	1.1
Renal ca. ACHN	11.9	Pancreatic ca. CAPAN2	1.0
Renal ca. UO-31	15.3	Pancreas Pool	28.7

Table AAD. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3229, Run 174442765	Tissue Name	Rel. Exp.(%) Ag3229, Run 174442765
Normal Colon	15.5	Kidney Margin (OD04348)	**100.0**
Colon cancer	31.9	Kidney malignant	10.7

643

(OD06064)		cancer (OD06204B)	
Colon Margin (OD06064)	20.6	Kidney normal adjacent tissue (OD06204E)	11.6
Colon cancer (OD06159)	6.0	Kidney Cancer (OD04450-01)	38.4
Colon Margin (OD06159)	12.7	Kidney Margin (OD04450-03)	17.4
Colon cancer (OD06297-04)	3.7	Kidney Cancer 8120613	0.0
Colon Margin (OD06297-05)	22.4	Kidney Margin 8120614	6.0
CC Gr.2 ascend colon (ODO3921)	6.5	Kidney Cancer 9010320	12.0
CC Margin (ODO3921)	10.5	Kidney Margin 9010321	9.9
Colon cancer metastasis (OD06104)	8.6	Kidney Cancer 8120607	47.3
Lung Margin (OD06104)	6.2	Kidney Margin 8120608	5.6
Colon mets to lung (OD04451-01)	31.0	Normal Uterus	48.3
Lung Margin (OD04451-02)	39.5	Uterine Cancer 064011	14.9
Normal Prostate	41.2	Normal Thyroid	2.6
Prostate Cancer (OD04410)	8.1	Thyroid Cancer 064010	4.3
Prostate Margin (OD04410)	10.6	Thyroid Cancer A302152	15.3
Normal Ovary	23.2	Thyroid Margin A302153	2.7
Ovarian cancer (OD06283-03)	7.2	Normal Breast	46.0
Ovarian Margin (OD06283-07)	17.8	Breast Cancer (OD04566)	5.9
Ovarian Cancer 064008	22.2	Breast Cancer 1024	27.4
Ovarian cancer (OD06145)	9.0	Breast Cancer (OD04590-01)	19.5
Ovarian Margin (OD06145)	13.4	Breast Cancer Mets (OD04590-03)	13.5
Ovarian cancer (OD06455-03)	6.3	Breast Cancer Metastasis (OD04655-05)	12.2
Ovarian Margin (OD06455-07)	12.9	Breast Cancer 064006	8.1

Normal Lung	14.5	Breast Cancer 9100266	3.0
Invasive poor diff. lung adeno (ODO4945-01	5.0	Breast Margin 9100265	3.4
Lung Margin (ODO4945-03)	37.4	Breast Cancer A209073	11.2
Lung Malignant Cancer (OD03126)	9.6	Breast Margin A2090734	61.1
Lung Margin (OD03126)	14.2	Breast cancer (OD06083)	4.7
Lung Cancer (OD05014A)	4.9	Breast cancer node metastasis (OD06083)	12.7
Lung Margin (OD05014B)	39.0	Normal Liver	2.8
Lung cancer (OD06081)	17.4	Liver Cancer 1026	13.6
Lung Margin (OD06081)	32.3	Liver Cancer 1025	12.9
Lung Cancer (OD04237-01)	4.2	Liver Cancer 6004-T	13.2
Lung Margin (OD04237-02)	24.7	Liver Tissue 6004-N	1.3
Ocular Melanoma Metastasis	12.9	Liver Cancer 6005-T	43.2
Ocular Melanoma Margin (Liver)	10.7	Liver Tissue 6005-N	4.8
Melanoma Metastasis	52.9	Liver Cancer 064003	39.5
Melanoma Margin (Lung)	21.0	Normal Bladder	9.3
Normal Kidney	4.7	Bladder Cancer 1023	5.8
Kidney Ca, Nuclear grade 2 (OD04338)	40.3	Bladder Cancer A302173	4.2
Kidney Margin (OD04338)	7.5	Normal Stomach	31.4
Kidney Ca Nuclear grade 1/2 (OD04339)	82.4	Gastric Cancer 9060397	1.2
Kidney Margin (OD04339)	13.2	Stomach Margin 9060396	7.1
Kidney Ca, Clear cell type (OD04340)	8.3	Gastric Cancer 9060395	7.4
Kidney Margin (OD04340)	24.7	Stomach Margin 9060394	10.9
Kidney Ca, Nuclear grade 3 (OD04348)	13.1	Gastric Cancer 064005	10.4

Table AAE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3229, Run 164389704	Tissue Name	Rel. Exp.(%) Ag3229, Run 164389704
Secondary Th1 act	3.4	HUVEC IL-1beta	20.0
Secondary Th2 act	4.8	HUVEC IFN gamma	32.5
Secondary Tr1 act	2.1	HUVEC TNF alpha + IFN gamma	26.6
Secondary Th1 rest	1.2	HUVEC TNF alpha + IL4	35.1
Secondary Th2 rest	2.0	HUVEC IL-11	17.6
Secondary Tr1 rest	4.5	Lung Microvascular EC none	34.2
Primary Th1 act	17.7	Lung Microvascular EC TNFalpha + IL-1beta	49.0
Primary Th2 act	5.3	Microvascular Dermal EC none	30.6
Primary Tr1 act	25.9	Microsvasular Dermal EC TNFalpha + IL-1beta	38.7
Primary Th1 rest	14.0	Bronchial epithelium TNFalpha + IL1beta	46.7
Primary Th2 rest	6.5	Small airway epithelium none	22.1
Primary Tr1 rest	22.1	Small airway epithelium TNFalpha + IL-1beta	97.9
CD45RA CD4 lymphocyte act	12.7	Coronery artery SMC rest	31.2
CD45RO CD4 lymphocyte act	6.6	Coronery artery SMC TNFalpha + IL-1beta	10.5
CD8 lymphocyte act	2.4	Astrocytes rest	7.5
Secondary CD8 lymphocyte rest	4.2	Astrocytes TNFalpha + IL-1beta	8.6
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.8
CD4 lymphocyte none	5.6	KU-812 (Basophil) PMA/ionomycin	2.9
2ry Th1/Th2/Tr1_anti-CD95 CH11	3.7	CCD1106 (Keratinocytes) none	6.2
LAK cells rest	5.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.0
LAK cells IL-2	3.0	Liver cirrhosis	15.5
LAK cells IL-2+IL-12	6.2	Lupus kidney	12.2
LAK cells IL-2+IFN gamma	10.7	NCI-H292 none	30.8
LAK cells IL-2+ IL-18	5.0	NCI-H292 IL-4	49.7

LAK cells PMA/ionomycin	4.0	NCI-H292 IL-9	43.5
NK Cells IL-2 rest	1.9	NCI-H292 IL-13	31.6
Two Way MLR 3 day	9.0	NCI-H292 IFN gamma	17.7
Two Way MLR 5 day	3.3	HPAEC none	18.0
Two Way MLR 7 day	1.2	HPAEC TNF alpha + IL-1 beta	58.2
PBMC rest	0.8	Lung fibroblast none	40.6
PBMC PWM	10.7	Lung fibroblast TNF alpha + IL-1 beta	11.0
PBMC PHA-L	10.2	Lung fibroblast IL-4	**100.0**
Ramos (B cell) none	0.0	Lung fibroblast IL-9	55.1
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	78.5
B lymphocytes PWM	23.3	Lung fibroblast IFN gamma	82.4
B lymphocytes CD40L and IL-4	18.6	Dermal fibroblast CCD1070 rest	45.4
EOL-1 dbcAMP	1.8	Dermal fibroblast CCD1070 TNF alpha	36.3
EOL-1 dbcAMP PMA/ionomycin	2.0	Dermal fibroblast CCD1070 IL-1 beta	23.8
Dendritic cells none	5.9	Dermal fibroblast IFN gamma	4.6
Dendritic cells LPS	8.0	Dermal fibroblast IL-4	16.6
Dendritic cells anti-CD40	3.3	IBD Colitis 2	6.2
Monocytes rest	4.2	IBD Crohn's	3.6
Monocytes LPS	0.9	Colon	30.1
Macrophages rest	3.1	Lung	19.9
Macrophages LPS	1.4	Thymus	20.6
HUVEC none	35.1	Kidney	20.9
HUVEC starved	58.2		

CNS_neurodegeneration_v1.0 Summary: Ag3229 - This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.4 Summary: Ag3229 - Highest levels of expression of this gene are seen in breast cancer cell line T47D (CT=28.5). Based on expression in this panel, this gene may be involved in gastric, brain, colon, renal, lung, breast, ovarian and prostate cancer as well as melanomas. Thus, expression of this gene could be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition using antibodies or small molecule drugs might be of use in the treatment of these cancers.

This gene product is also expressed in adipose, pancreas, adrenal, thyroid, pituitary, skeletal muscle, and heart. This widespread expression in tissues with metabolic function suggests that this gene product may be important for the pathogenesis, diagnosis, and/or treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

In addition, this gene is expressed at low to moderate levels in all regions of the CNS examined. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 2.2 Summary: Ag3229 Highest expression of the CG57411-01 gene is seen in the kidney (CT=32.2). In addition, significant levels of expression are seen in samples derived from normal lung and breast. Expression in these normal tissues is also higher than in the corresponding malignant tissue. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of lung, breast and kidney cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of lung, breast and kidney cancer.

Panel 4D Summary: Ag3229 Highest expression of the CG57411-01 gene is seen in IL-4 treated lung fibroblasts (CT=31.3). Significant levels of expression are seen in activated-NCI-H292 mucoepidermoid cells as well as untreated NCI-H292 cells. Moderate expression is also detected in IL-9, IL-13 and IFN gamma activated lung fibroblasts, human pulmonary aortic endothelial cells (treated and untreated), small airway epithelium (treated and untreated), treated bronchial epithelium and lung microvascular endothelial cells (treated and untreated). The expression of this gene in cells derived from or within the lung suggests that this gene may be involved in normal conditions as well as pathological

648

and inflammatory lung disorders that include chronic obstructive pulmonary disease, asthma, allergy and emphysema. Moderate/low expression of this gene is also detected in treated and untreated HUVECs (endothelial cells) and coronary artery smooth muscle cells (treated and untreated) and normal tissues that include lung, colon, thymus and kidney.

Expression in the various immune cell types and tissue samples suggests that therapeutic modulation of this gene product may ameliorate symptoms associated with infectious conditions as well as inflammatory and autoimmune disorders that include psoriasis, allergy, asthma, inflammatory bowel disease, rheumatoid arthritis and osteoarthritis.

AB. CG57399-01 and CG57399-03: PHOSPHOLIPASE ADRAB-B PRECURSOR

Expression of gene CG57399-01 and variant CG57399-03 was assessed using the primer-probe sets Ag3952 and Ag3226, described in Tables ABA and ABB. Results of the RTQ-PCR runs are shown in Tables ABC and ABD.

Table ABA. Probe Name Ag3952

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctgtgtccctgtgtcctgaa-3'	20	1633	475
Probe	TET-5'-tcaacagaacttgctaccctcatcga-3'-TAMRA	26	1666	476
Reverse	5'-gtgggtcttctcctgaaacttc-3'	22	1701	477

Table ABB. Probe Name Ag3226

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gatgatcctcaggtcactgtgt-3'	22	1617	478
Probe	TET-5'-ccctgtgtcctgaagtttgatgataactca-3'-TAMRA	30	1639	479
Reverse	5'-tcgatgagggtagcaagttct-3'	21	1671	480

Table ABC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3952, Run 213856126	Tissue Name	Rel. Exp.(%) Ag3952, Run 213856126
Adipose	9.0	Renal ca. TK-10	15.0
Melanoma* Hs688(A).T	3.0	Bladder	22.7
Melanoma* Hs688(B).T	3.4	Gastric ca. (liver met.) NCI-N87	13.0

Melanoma* M14	0.9	Gastric ca. KATO III	75.3
Melanoma* LOXIMVI	11.7	Colon ca. SW-948	4.3
Melanoma* SK-MEL-5	1.5	Colon ca. SW480	97.3
Squamous cell carcinoma SCC-4	8.7	Colon ca.* (SW480 met) SW620	4.4
Testis Pool	12.8	Colon ca. HT29	0.4
Prostate ca.* (bone met) PC-3	10.5	Colon ca. HCT-116	1.2
Prostate Pool	12.9	Colon ca. CaCo-2	60.7
Placenta	5.1	Colon cancer tissue	28.7
Uterus Pool	6.5	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	7.3	Colon ca. Colo-205	0.9
Ovarian ca. SK-OV-3	26.4	Colon ca. SW-48	26.1
Ovarian ca. OVCAR-4	1.9	Colon Pool	18.8
Ovarian ca. OVCAR-5	6.7	Small Intestine Pool	5.3
Ovarian ca. IGROV-1	9.2	Stomach Pool	7.9
Ovarian ca. OVCAR-8	4.2	Bone Marrow Pool	8.4
Ovary	10.0	Fetal Heart	1.2
Breast ca. MCF-7	0.4	Heart Pool	5.7
Breast ca. MDA-MB-231	92.0	Lymph Node Pool	32.1
Breast ca. BT 549	5.5	Fetal Skeletal Muscle	1.2
Breast ca. T47D	2.5	Skeletal Muscle Pool	4.7
Breast ca. MDA-N	1.6	Spleen Pool	18.2
Breast Pool	19.6	Thymus Pool	19.3
Trachea	10.3	CNS cancer (glio/astro) U87-MG	38.2
Lung	1.2	CNS cancer (glio/astro) U-118-MG	12.2
Fetal Lung	8.3	CNS cancer (neuro;met) SK-N-AS	0.9
Lung ca. NCI-N417	0.9	CNS cancer (astro) SF-539	7.6
Lung ca. LX-1	27.2	CNS cancer (astro) SNB-75	17.1
Lung ca. NCI-H146	10.7	CNS cancer (glio) SNB-19	6.8

650

Lung ca. SHP-77	47.3	CNS cancer (glio) SF-295	5.7
Lung ca. A549	5.1	Brain (Amygdala) Pool	7.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	3.2
Lung ca. NCI-H23	4.1	Brain (fetal)	19.3
Lung ca. NCI-H460	0.5	Brain (Hippocampus) Pool	13.1
Lung ca. HOP-62	2.7	Cerebral Cortex Pool	14.8
Lung ca. NCI-H522	1.3	Brain (Substantia nigra) Pool	6.3
Liver	0.0	Brain (Thalamus) Pool	15.2
Fetal Liver	1.7	Brain (whole)	10.4
Liver ca. HepG2	0.5	Spinal Cord Pool	5.3
Kidney Pool	21.2	Adrenal Gland	**100.0**
Fetal Kidney	1.6	Pituitary gland Pool	4.3
Renal ca. 786-0	1.7	Salivary Gland	3.4
Renal ca. A498	1.3	Thyroid (female)	14.5
Renal ca. ACHN	4.3	Pancreatic ca. CAPAN2	1.7
Renal ca. UO-31	17.4	Pancreas Pool	24.5

Table ABD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3226, Run 167994701	Tissue Name	Rel. Exp.(%) Ag3226, Run 167994701
Liver adenocarcinoma	2.5	Kidney (fetal)	16.3
Pancreas	0.0	Renal ca. 786-0	0.9
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	1.4
Adrenal gland	19.6	Renal ca. RXF 393	3.4
Thyroid	16.3	Renal ca. ACHN	1.4
Salivary gland	0.0	Renal ca. UO-31	2.8
Pituitary gland	1.9	Renal ca. TK-10	4.4
Brain (fetal)	25.5	Liver	0.0
Brain (whole)	4.6	Liver (fetal)	1.1
Brain (amygdala)	6.7	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	1.6	Lung	8.8
Brain (hippocampus)	22.2	Lung (fetal)	1.7
Brain (substantia nigra)	3.1	Lung ca. (small cell) LX-1	18.6
Brain (thalamus)	3.2	Lung ca. (small cell) NCI-H69	4.2

Cerebral Cortex	26.2	Lung ca. (s.cell var.) SHP-77	**100.0**
Spinal cord	3.1	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	7.5	Lung ca. (non-sm. cell) A549	6.7
glio/astro U-118-MG	4.2	Lung ca. (non-s.cell) NCI-H23	5.7
astrocytoma SW1783	1.2	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.9
astrocytoma SNB-75	4.3	Lung ca. (squam.) NCI-H596	3.7
glioma SNB-19	6.0	Mammary gland	6.3
glioma U251	14.1	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	45.4
Heart (fetal)	1.4	Breast ca.* (pl.ef) T47D	4.3
Heart	1.0	Breast ca. BT-549	7.1
Skeletal muscle (fetal)	0.7	Breast ca. MDA-N	0.0
Skeletal muscle	3.2	Ovary	10.9
Bone marrow	3.1	Ovarian ca. OVCAR-3	0.0
Thymus	5.7	Ovarian ca. OVCAR-4	2.4
Spleen	7.2	Ovarian ca. OVCAR-5	5.2
Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	4.8	Ovarian ca. IGROV-1	0.0
Stomach	5.1	Ovarian ca.* (ascites) SK-OV-3	3.0
Small intestine	1.5	Uterus	5.8
Colon ca. SW480	33.2	Placenta	0.0
Colon ca.* SW620(SW480 met)	8.8	Prostate	1.6
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	2.6
Colon ca. HCT-116	0.0	Testis	7.4

652

Colon ca. CaCo-2	35.4	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	24.5	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	15.7	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	6.4	Melanoma M14	0.0
Bladder	14.6	Melanoma LOX IMVI	0.0
Trachea	4.4	Melanoma* (met) SK-MEL-5	0.0
Kidney	2.4	Adipose	17.3

General_screening_panel_v1.4 Summary: Ag3952 Highest expression of this gene is seen in the adrenal gland (CT=29). Thus, this gene product may be a treatment for Addison's disease and other adrenalopathies. This gene also has low levels of expression in adipose, heart, skeletal muscle, pituitary, thyroid, and pancreas. Therapeutic modulation of this gene product may be important for the diagnosis or treatment of endocrine or metabolic disease, including Types 1 and 2 diabetes, obesity and pancreatitis.

Expression of this gene is also seen in sample derived from colon, gastric, lung and breast cancers. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of colon, gastric, lung and breast cancers.

Low but significant levels of expression are also seen for all regions of the CNS examined. Thus, this gene product may be useful for treatment of CNS disorders such as Alzheimer's disease, Parkinson's disease, stroke, epilepsy, schizophrenia and multiple sclerosis.

Panel 1.3D Summary: Ag3952 Highest expression of the CG57399-01 gene is seen in a lung cancer cell line (CT=32.5). Low but significant expression is also seen in cell lines derived from breast and colon cancers. Overall, expression is consistent with expression seen in Panel 1.4. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of colon, gastric, lung and breast cancers.

Among metabolic tissues, significant levels of expression are seen in adipose and the adrenal gland. Thus, this gene product may be useful for treatment of obesity, Addison's disease and other adrenalopathies.

In addition, this gene is expressed in the hippocampus, and cerebral cortex. Both these regions of the brain undergo degeneration in Alzheimer's disease. Thus, therapeutic modulation of the expression or function of this gene may be effective in the treatment of this disease or any other neurodegenerative disorders.

AC. CG57399-02: PHOSPHOLIPASE ADRAB-B PRECURSOR

Expression of gene CG57399-02 was assessed using the primer-probe set Ag3952, described in Table ACA. Results of the RTQ-PCR runs are shown in Table ACB. Please note that this gene represents a variant of CG57399-01. This sequence however, only corresponds to probe and primer set Ag3952.

Table ACA. Probe Name Ag3952

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctgtgtccctgtgtcctgaa-3'	20	578	481
Probe	TET-5'-tcaacagaacttgctaccctcatcga-3'-TAMRA	26	611	482
Reverse	5'-gtgggtcttctcctgaaacttc-3'	22	646	483

Table ACB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3952, Run 213856126	Tissue Name	Rel. Exp.(%) Ag3952, Run 213856126
Adipose	9.0	Renal ca. TK-10	15.0
Melanoma* Hs688(A).T	3.0	Bladder	22.7
Melanoma* Hs688(B).T	3.4	Gastric ca. (liver met.) NCI-N87	13.0
Melanoma* M14	0.9	Gastric ca. KATO III	75.3
Melanoma* LOXIMVI	11.7	Colon ca. SW-948	4.3
Melanoma* SK-MEL-5	1.5	Colon ca. SW480	97.3
Squamous cell carcinoma SCC-4	8.7	Colon ca.* (SW480 met) SW620	4.4

Testis Pool	12.8	Colon ca. HT29	0.4
Prostate ca.* (bone met) PC-3	10.5	Colon ca. HCT-116	1.2
Prostate Pool	12.9	Colon ca. CaCo-2	60.7
Placenta	5.1	Colon cancer tissue	28.7
Uterus Pool	6.5	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	7.3	Colon ca. Colo-205	0.9
Ovarian ca. SK-OV-3	26.4	Colon ca. SW-48	26.1
Ovarian ca. OVCAR-4	1.9	Colon Pool	18.8
Ovarian ca. OVCAR-5	6.7	Small Intestine Pool	5.3
Ovarian ca. IGROV-1	9.2	Stomach Pool	7.9
Ovarian ca. OVCAR-8	4.2	Bone Marrow Pool	8.4
Ovary	10.0	Fetal Heart	1.2
Breast ca. MCF-7	0.4	Heart Pool	5.7
Breast ca. MDA-MB-231	92.0	Lymph Node Pool	32.1
Breast ca. BT 549	5.5	Fetal Skeletal Muscle	1.2
Breast ca. T47D	2.5	Skeletal Muscle Pool	4.7
Breast ca. MDA-N	1.6	Spleen Pool	18.2
Breast Pool	19.6	Thymus Pool	19.3
Trachea	10.3	CNS cancer (glio/astro) U87-MG	38.2
Lung	1.2	CNS cancer (glio/astro) U-118-MG	12.2
Fetal Lung	8.3	CNS cancer (neuro;met) SK-N-AS	0.9
Lung ca. NCI-N417	0.9	CNS cancer (astro) SF-539	7.6
Lung ca. LX-1	27.2	CNS cancer (astro) SNB-75	17.1
Lung ca. NCI-H146	10.7	CNS cancer (glio) SNB-19	6.8
Lung ca. SHP-77	47.3	CNS cancer (glio) SF-295	5.7
Lung ca. A549	5.1	Brain (Amygdala) Pool	7.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	3.2
Lung ca. NCI-H23	4.1	Brain (fetal)	19.3
Lung ca. NCI-H460	0.5	Brain (Hippocampus)	13.1

		Pool	
Lung ca. HOP-62	2.7	Cerebral Cortex Pool	14.8
Lung ca. NCI-H522	1.3	Brain (Substantia nigra) Pool	6.3
Liver	0.0	Brain (Thalamus) Pool	15.2
Fetal Liver	1.7	Brain (whole)	10.4
Liver ca. HepG2	0.5	Spinal Cord Pool	5.3
Kidney Pool	21.2	Adrenal Gland	**100.0**
Fetal Kidney	1.6	Pituitary gland Pool	4.3
Renal ca. 786-0	1.7	Salivary Gland	3.4
Renal ca. A498	1.3	Thyroid (female)	14.5
Renal ca. ACHN	4.3	Pancreatic ca. CAPAN2	1.7
Renal ca. UO-31	17.4	Pancreas Pool	24.5

General_screening_panel_v1.4 Summary: Ag3952 Highest expression of this gene is seen in the adrenal gland (CT=29). Thus, this gene product may be a treatment for Addison's disease and other adrenalopathies. This gene also has low levels of expression in adipose, heart, skeletal muscle, pituitary, thyroid, and pancreas. Therapeutic modulation of this gene product may be important for the diagnosis or treatment of endocrine or metabolic disease, including Types 1 and 2 diabetes, obesity and pancreatitis.

Expression of this gene is also seen in cell line samples derived from colon, gastric, lung and breast cancers. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of colon, gastric, lung and breast cancers.

Low but significant levels of expression are also seen for all regions of the CNS examined. Thus, this gene product may be useful for treatment of CNS disorders such as Alzheimer's disease, Parkinson's disease, stroke, epilepsy, schizophrenia and multiple sclerosis.

AD. CG59311-01: ACYL-COENZYME A THIOESTER HYDROLASE bp.

Expression of gene CG59311-01, splice variant CG59311-02, and full length clone CG59311-03, was assessed using the primer-probe set Ag3541, described in Table ADA. Results of the RTQ-PCR runs are shown in Tables ADB and ADC.

656

Table ADA. Probe Name Ag3541

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctcactcaaaggcacaggtaga-3'	22	1199	484
Probe	TET-5'-tggcagcaaattcaaactttcttcca-3'-TAMRA	26	1225	485
Reverse	5'-tttgctgtgcttgacagatttt-3'	22	1269	486

Table ADB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3541, Run 217049294	Tissue Name	Rel. Exp.(%) Ag3541, Run 217049294
Adipose	0.0	Renal ca. TK-10	6.0
Melanoma* Hs688(A).T	0.7	Bladder	3.7
Melanoma* Hs688(B).T	1.6	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	2.7
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	5.4
Testis Pool	3.1	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	1.4	Colon ca. HCT-116	0.6
Prostate Pool	2.3	Colon ca. CaCo-2	0.6
Placenta	0.5	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	2.9	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.9	Colon Pool	4.6
Ovarian ca. OVCAR-5	27.0	Small Intestine Pool	6.6
Ovarian ca. IGROV-1	0.0	Stomach Pool	3.1
Ovarian ca. OVCAR-8	1.8	Bone Marrow Pool	1.4
Ovary	2.5	Fetal Heart	9.2

657

Breast ca. MCF-7	2.4	Heart Pool	3.4
Breast ca. MDA-MB-231	8.0	Lymph Node Pool	3.9
Breast ca. BT 549	4.9	Fetal Skeletal Muscle	4.9
Breast ca. T47D	52.9	Skeletal Muscle Pool	13.5
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	6.7	Thymus Pool	4.7
Trachea	0.9	CNS cancer (glio/astro) U87-MG	0.9
Lung	1.7	CNS cancer (glio/astro) U-118-MG	12.1
Fetal Lung	2.2	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	4.2	CNS cancer (astro) SNB-75	5.2
Lung ca. NCI-H146	2.1	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	6.7	CNS cancer (glio) SF-295	0.7
Lung ca. A549	0.0	Brain (Amygdala) Pool	4.2
Lung ca. NCI-H526	0.0	Brain (cerebellum)	**100.0**
Lung ca. NCI-H23	10.2	Brain (fetal)	14.7
Lung ca. NCI-H460	3.4	Brain (Hippocampus) Pool	9.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	9.7
Lung ca. NCI-H522	8.5	Brain (Substantia nigra) Pool	3.5
Liver	0.5	Brain (Thalamus) Pool	10.5
Fetal Liver	1.5	Brain (whole)	12.9
Liver ca. HepG2	0.5	Spinal Cord Pool	7.6
Kidney Pool	9.8	Adrenal Gland	10.2
Fetal Kidney	7.9	Pituitary gland Pool	3.1
Renal ca. 786-0	0.0	Salivary Gland	1.7
Renal ca. A498	0.0	Thyroid (female)	0.9
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	2.8
Renal ca. UO-31	3.3	Pancreas Pool	6.6

Table ADC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3541, Run	Tissue Name	Rel. Exp.(%) Ag3541, Run

	166447041		166447041
Secondary Th1 act	2.7	HUVEC IL-1beta	0.0
Secondary Th2 act	4.1	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	2.1
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	2.7	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	2.3
CD45RO CD4 lymphocyte act	1.7	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	1.8
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	4.2
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	1.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	2.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	9.8
LAK cells IL-2	0.0	Liver cirrhosis	22.2
LAK cells IL-2+IL-12	0.0	Lupus kidney	18.4
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	10.4
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	7.2
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	15.2

659

NK Cells IL-2 rest	1.7	NCI-H292 IL-13	3.1
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	7.3
Two Way MLR 5 day	5.3	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	1.7
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	5.7
PBMC PHA-L	2.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	2.2	Lung fibroblast IL-13	3.2
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	2.9
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	2.9
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	3.5	Dermal fibroblast IL-4	1.5
Dendritic cells anti-CD40	0.0	IBD Colitis 2	5.4
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	14.1
Macrophages rest	4.5	Lung	0.0
Macrophages LPS	2.1	Thymus	**100.0**
HUVEC none	0.0	Kidney	2.3
HUVEC starved	2.5		

CNS_neurodegeneration_v1.0 Summary: Ag3541 - Expression of this gene is low/undetectable (CTs > 34.5) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3541 Significant expression of this gene is seen only in cerebellum, fetal brain, the breast cancer cell line T47D, and ovarian cancer cell line OVCAR-5 (CTs=32-35). Therefore, expression of this gene can be used to differentiate between these samples and others on this panel.

Panel 4D Summary: Ag3541 - There is significant expression of this gene only in thymus (CT=33.8). Therefore, expression of this gene may be used to identify thymic tissue.

660

Furthermore, drugs that inhibit the function of this protein may regulate T cell development in the thymus and reduce or eliminate the symptoms of T cell mediated autoimmune or inflammatory diseases, including asthma, allergies, inflammatory bowel disease, lupus erythematosus, or rheumatoid arthritis. Additionally, therapeutics designed against this putative protein may disrupt T cell development in the thymus and function as an immunosuppresant for tissue transplant.

AE. CG59309-01: ACYL-COENZYME A THIOESTER HYDROLASE

Expression of gene CG59309-01 was assessed using the primer-probe set Ag3540, described in Table AEA. Results of the RTQ-PCR runs are shown in Tables AEB, AEC, AED and AEE.

Table AEA. Probe Name Ag3540

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ccacgttggctctagcttatta-3'	22	649	487
Probe	TET-5'-tgaagatctccccaataacatggaca-3'-TAMRA	26	677	488
Reverse	5'-ttcgaagtactccagggatatg-3'	22	704	489

Table AEB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3540, Run 210638385	Tissue Name	Rel. Exp.(%) Ag3540, Run 210638385
AD 1 Hippo	13.7	Control (Path) 3 Temporal Ctx	8.2
AD 2 Hippo	26.2	Control (Path) 4 Temporal Ctx	34.2
AD 3 Hippo	13.1	AD 1 Occipital Ctx	23.2
AD 4 Hippo	3.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	30.4	AD 3 Occipital Ctx	7.8
AD 6 Hippo	55.9	AD 4 Occipital Ctx	15.0
Control 2 Hippo	24.0	AD 5 Occipital Ctx	8.1
Control 4 Hippo	4.5	AD 6 Occipital Ctx	76.3

Control (Path) 3 Hippo	6.2	Control 1 Occipital Ctx	3.6
AD 1 Temporal Ctx	11.0	Control 2 Occipital Ctx	96.6
AD 2 Temporal Ctx	19.5	Control 3 Occipital Ctx	36.3
AD 3 Temporal Ctx	4.8	Control 4 Occipital Ctx	3.9
AD 4 Temporal Ctx	15.6	Control (Path) 1 Occipital Ctx	**100.0**
AD 5 Inf Temporal Ctx	36.9	Control (Path) 2 Occipital Ctx	7.6
AD 5 SupTemporal Ctx	27.4	Control (Path) 3 Occipital Ctx	1.6
AD 6 Inf Temporal Ctx	47.3	Control (Path) 4 Occipital Ctx	16.6
AD 6 Sup Temporal Ctx	64.2	Control 1 Parietal Ctx	8.7
Control 1 Temporal Ctx	7.0	Control 2 Parietal Ctx	20.7
Control 2 Temporal Ctx	53.2	Control 3 Parietal Ctx	27.2
Control 3 Temporal Ctx	19.9	Control (Path) 1 Parietal Ctx	88.9
Control 4 Temporal Ctx	10.5	Control (Path) 2 Parietal Ctx	10.8
Control (Path) 1 Temporal Ctx	68.3	Control (Path) 3 Parietal Ctx	10.1
Control (Path) 2 Temporal Ctx	25.3	Control (Path) 4 Parietal Ctx	47.6

Table AEC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3540, Run 217049291	Tissue Name	Rel. Exp.(%) Ag3540, Run 217049291
Adipose	1.3	Renal ca. TK-10	0.1
Melanoma* Hs688(A).T	0.7	Bladder	1.1
Melanoma* Hs688(B).T	0.5	Gastric ca. (liver met.) NCI-N87	5.6
Melanoma* M14	0.2	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-	0.0	Colon ca. SW480	10.3

MEL-5			
Squamous cell carcinoma SCC-4	0.3	Colon ca.* (SW480 met) SW620	2.8
Testis Pool	0.3	Colon ca. HT29	0.8
Prostate ca.* (bone met) PC-3	0.8	Colon ca. HCT-116	0.0
Prostate Pool	0.3	Colon ca. CaCo-2	3.5
Placenta	1.4	Colon cancer tissue	1.4
Uterus Pool	0.1	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	1.6	Colon ca. Colo-205	3.3
Ovarian ca. SK-OV-3	3.6	Colon ca. SW-48	1.7
Ovarian ca. OVCAR-4	0.4	Colon Pool	0.2
Ovarian ca. OVCAR-5	23.7	Small Intestine Pool	0.3
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.1
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.2
Ovary	0.1	Fetal Heart	0.4
Breast ca. MCF-7	0.0	Heart Pool	0.2
Breast ca. MDA-MB-231	2.5	Lymph Node Pool	0.3
Breast ca. BT 549	3.0	Fetal Skeletal Muscle	0.1
Breast ca. T47D	**100.0**	Skeletal Muscle Pool	0.4
Breast ca. MDA-N	0.0	Spleen Pool	0.2
Breast Pool	0.3	Thymus Pool	0.3
Trachea	0.4	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.3
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	1.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.6
Lung ca. LX-1	3.5	CNS cancer (astro) SNB-75	3.1
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.1	CNS cancer (glio) SF-295	0.2
Lung ca. A549	1.4	Brain (Amygdala) Pool	0.7

Lung ca. NCI-H526	0.7	Brain (cerebellum)	2.1
Lung ca. NCI-H23	1.3	Brain (fetal)	0.5
Lung ca. NCI-H460	0.8	Brain (Hippocampus) Pool	1.0
Lung ca. HOP-62	1.2	Cerebral Cortex Pool	0.9
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	1.3
Liver	2.6	Brain (Thalamus) Pool	1.1
Fetal Liver	0.8	Brain (whole)	1.4
Liver ca. HepG2	0.1	Spinal Cord Pool	0.5
Kidney Pool	0.7	Adrenal Gland	0.8
Fetal Kidney	0.6	Pituitary gland Pool	0.1
Renal ca. 786-0	0.0	Salivary Gland	0.2
Renal ca. A498	0.0	Thyroid (female)	0.7
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	9.4
Renal ca. UO-31	1.1	Pancreas Pool	0.9

Table AED. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3540, Run 166447040	Tissue Name	Rel. Exp.(%) Ag3540, Run 166447040
Secondary Th1 act	4.8	HUVEC IL-1beta	1.7
Secondary Th2 act	10.2	HUVEC IFN gamma	0.9
Secondary Tr1 act	12.9	HUVEC TNF alpha + IFN gamma	1.5
Secondary Th1 rest	2.1	HUVEC TNF alpha + IL4	0.8
Secondary Th2 rest	1.4	HUVEC IL-11	1.5
Secondary Tr1 rest	1.6	Lung Microvascular EC none	0.6
Primary Th1 act	4.7	Lung Microvascular EC TNFalpha + IL-1beta	0.8
Primary Th2 act	6.8	Microvascular Dermal EC none	1.5
Primary Tr1 act	7.3	Microsvasular Dermal EC TNFalpha + IL-1beta	0.8
Primary Th1 rest	6.6	Bronchial epithelium TNFalpha + IL1beta	1.3
Primary Th2 rest	2.6	Small airway epithelium none	0.6
Primary Tr1 rest	4.2	Small airway epithelium TNFalpha + IL-1beta	0.0

CD45RA CD4 lymphocyte act	4.1	Coronery artery SMC rest	0.9
CD45RO CD4 lymphocyte act	10.9	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	6.6	Astrocytes rest	2.6
Secondary CD8 lymphocyte rest	17.0	Astrocytes TNFalpha + IL-1beta	2.1
Secondary CD8 lymphocyte act	6.0	KU-812 (Basophil) rest	2.2
CD4 lymphocyte none	2.0	KU-812 (Basophil) PMA/ionomycin	10.2
2ry Th1/Th2/Tr1_anti-CD95 CH11	2.4	CCD1106 (Keratinocytes) none	6.8
LAK cells rest	2.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	25.7
LAK cells IL-2	16.2	Liver cirrhosis	12.0
LAK cells IL-2+IL-12	12.8	Lupus kidney	5.1
LAK cells IL-2+IFN gamma	15.6	NCI-H292 none	44.8
LAK cells IL-2+ IL-18	7.4	NCI-H292 IL-4	37.6
LAK cells PMA/ionomycin	3.4	NCI-H292 IL-9	41.2
NK Cells IL-2 rest	9.0	NCI-H292 IL-13	19.8
Two Way MLR 3 day	10.5	NCI-H292 IFN gamma	30.1
Two Way MLR 5 day	7.2	HPAEC none	1.2
Two Way MLR 7 day	8.9	HPAEC TNF alpha + IL-1 beta	3.3
PBMC rest	0.5	Lung fibroblast none	0.9
PBMC PWM	3.8	Lung fibroblast TNF alpha + IL-1 beta	0.7
PBMC PHA-L	1.0	Lung fibroblast IL-4	0.5
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.9
B lymphocytes PWM	10.3	Lung fibroblast IFN gamma	1.2
B lymphocytes CD40L and IL-4	3.8	Dermal fibroblast CCD1070 rest	1.1
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	18.9
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	1.9
Dendritic cells none	14.9	Dermal fibroblast IFN gamma	0.0

665

Dendritic cells LPS	8.9	Dermal fibroblast IL-4	1.5
Dendritic cells anti-CD40	7.9	IBD Colitis 2	2.9
Monocytes rest	0.0	IBD Crohn's	1.9
Monocytes LPS	0.6	Colon	82.9
Macrophages rest	40.3	Lung	9.7
Macrophages LPS	6.1	Thymus	**100.0**
HUVEC none	1.1	Kidney	1.8
HUVEC starved	1.4		

Table AEE. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag3540, Run 242386396	Tissue Name	Rel. Exp.(%) Ag3540, Run 242386396
97457_Patient-02go_adipose	3.3	94709_Donor 2 AM - A_adipose	9.1
97476_Patient-07sk_skeletal muscle	0.8	94710_Donor 2 AM - B_adipose	1.6
97477_Patient-07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	1.4
97478_Patient-07pl_placenta	12.9	94712_Donor 2 AD - A_adipose	2.8
99167_Bayer Patient 1	15.5	94713_Donor 2 AD - B_adipose	5.8
97482_Patient-08ut_uterus	3.4	94714_Donor 2 AD - C_adipose	4.2
97483_Patient-08pl_placenta	3.4	94742_Donor 3 U - A_Mesenchymal Stem Cells	3.0
97486_Patient-09sk_skeletal muscle	**100.0**	94743_Donor 3 U - B_Mesenchymal Stem Cells	1.1
97487_Patient-09ut_uterus	1.6	94730_Donor 3 AM - A_adipose	4.3
97488_Patient-09pl_placenta	2.6	94731_Donor 3 AM - B_adipose	2.0
97492_Patient-10ut_uterus	3.1	94732_Donor 3 AM - C_adipose	2.0
97493_Patient-10pl_placenta	23.2	94733_Donor 3 AD - A_adipose	10.7
97495_Patient-11go_adipose	0.8	94734_Donor 3 AD - B_adipose	3.0
97496_Patient-11sk_skeletal muscle	0.0	94735_Donor 3 AD - C_adipose	4.0
97497_Patient-11ut_uterus	2.5	77138_Liver_HepG2untreated	0.7

666

97498_Patient-11pl_placenta	6.7	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	6.5	81735_Small Intestine	4.8
97501_Patient-12sk_skeletal muscle	4.5	72409_Kidney_Proximal Convoluted Tubule	0.7
97502_Patient-12ut_uterus	6.7	82685_Small intestine_Duodenum	3.6
97503_Patient-12pl_placenta	2.4	90650_Adrenal_Adrenocortical adenoma	0.6
94721_Donor 2 U - A_Mesenchymal Stem Cells	2.2	72410_Kidney_HRCE	8.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.6	72411_Kidney_HRE	8.5
94723_Donor 2 U - C_Mesenchymal Stem Cells	3.1	73139_Uterus_Uterine smooth muscle cells	0.0

CNS_neurodegeneration_v1.0 Summary: Ag3540 - This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment.

General_screening_panel_v1.4 Summary: Ag3540 This gene is most highly expressed in a breast cancer cell line (CT=27.1). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of breast cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of breast cancer.

Among metabolic tissues, this gene, an acyl coA thioesterase homolog, has a low level of expression in adipose, adult and fetal liver, adrenal, thyroid and pancreas. Acyl CoA thioesterases have multiple roles in lipid homeostasis. Therefore, therapeutic modulation of this gene product may be a treatment for endocrine and metabolic disease, including Types 1 and 2 diabetes and obesity.

In addition, this gene is expressed in all CNS regions examined. Thus, therapeutic modulation of the expression or function of this gene may be effective in the treatment of

667

neurologic disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, stroke, schizophrenia and multiple sclerosis.

References:

1. Hunt MC, Alexson SE. The role Acyl-CoA thioesterases play in mediating
intracellular lipid metabolism. Prog Lipid Res. 2002 Mar;41(2):99-130.

2. Hunt MC, Nousiainen SE, Huttunen MK, Orii KE, Svensson LT, Alexson SE.
Peroxisome proliferator-induced long chain acyl-CoA thioesterases comprise a highly
conserved novel multi-gene family involved in lipid metabolism. J Biol Chem. 1999 Nov
26;274(48):34317-26.

Panel 4D Summary: Ag3540 Highest expression of the CG59309-01 gene is seen in the thymus and colon (CTs=31.5). Significant levels of expression are also seen in a cluter of treated and untreated samples derived from the NCI-H292 mucoepidermoid cell line. Thus, expression of this gene could be used as a marker for thymus and colon. Furthermore, therapeutic modulation of the expression or function of this gene may regulate T cell development in the thymus and reduce or eliminate the symptoms of T cell mediated autoimmune or inflammatory diseases, including asthma, allergies, inflammatory bowel disease, lupus erythematosus, or rheumatoid arthritis. Additionally, small molecule or antibody therapeutics designed against this putative protein may disrupt T cell development in the thymus and function as an immunosuppresant for tissue transplant.

Panel 5 Islet Summary: Ag3540 This gene has moderate expression in skeletal muscle, (highest expression CT=30.5). Acyl CoA thioesterases function in peroxisomal fatty acid oxidation. Therefore, therapeutic modulation of this homolog may increase fatty acid oxidation in muscle and be a treatment for Type 2 diabetes and obesity.

References:

1. Hunt MC, Solaas K, Kase BF, Alexson SE. Characterization of an acyl-coA
thioesterase that functions as a major regulator of peroxisomal lipid metabolism. J Biol
Chem. 2002 Jan 11;277(2):1128-38.

AF. CG57364-01: CG6896

Expression of gene CG57364-01 was assessed using the primer-probe sets Ag3218 and Ag3378, described in Tables AFA and AFB. Results of the RTQ-PCR runs are shown in Tables AFC, AFD, AFE and AFF.

Table AFA. Probe Name Ag3218

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctcctgaagcaggtcctctt-3'	20	249	490
Probe	TET-5'-cctcccagtgttgtccttctggagg-3'-TAMRA	25	270	491
Reverse	5'-gacttcttccaggtcatttcg-3'	21	303	492

Table AFB. Probe Name Ag3378

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctcctgaagcaggtcctctt-3'	20	249	493
Probe	TET-5'-cctcccagtgttgtccttctggagg-3'-TAMRA	25	270	494
Reverse	5'-gacttcttccaggtcatttcg-3'	21	303	495

Table AFC. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3218, Run 209861784	Rel. Exp.(%) Ag3378, Run 210154573	Tissue Name	Rel. Exp.(%) Ag3218, Run 209861784	Rel. Exp.(%) Ag3378, Run 210154573
AD 1 Hippo	37.6	30.4	Control (Path) 3 Temporal Ctx	17.6	16.7
AD 2 Hippo	31.0	37.6	Control (Path) 4 Temporal Ctx	37.6	31.2
AD 3 Hippo	34.2	21.5	AD 1 Occipital Ctx	56.3	40.3
AD 4 Hippo	40.6	25.3	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	**100.0**	69.3	AD 3	43.2	24.1

669

			Occipital Ctx		
AD 6 Hippo	62.9	55.9	AD 4 Occipital Ctx	80.1	24.3
Control 2 Hippo	55.1	52.9	AD 5 Occipital Ctx	17.9	25.2
Control 4 Hippo	35.4	39.5	AD 6 Occipital Ctx	66.9	55.5
Control (Path) 3 Hippo	22.8	26.8	Control 1 Occipital Ctx	27.9	17.4
AD 1 Temporal Ctx	40.3	28.3	Control 2 Occipital Ctx	94.0	64.6
AD 2 Temporal Ctx	83.5	94.6	Control 3 Occipital Ctx	43.5	40.6
AD 3 Temporal Ctx	30.8	24.5	Control 4 Occipital Ctx	20.3	22.5
AD 4 Temporal Ctx	61.1	26.8	Control (Path) 1 Occipital Ctx	79.6	51.4
AD 5 Inf Temporal Ctx	84.7	**100.0**	Control (Path) 2 Occipital Ctx	34.4	24.7
AD 5 SupTemporal Ctx	55.9	39.8	Control (Path) 3 Occipital Ctx	25.2	16.2
AD 6 Inf Temporal Ctx	47.0	46.0	Control (Path) 4 Occipital Ctx	76.3	45.1
AD 6 Sup Temporal Ctx	63.7	41.2	Control 1 Parietal Ctx	31.0	21.9
Control 1 Temporal Ctx	32.8	18.0	Control 2 Parietal Ctx	67.4	45.1
Control 2 Temporal Ctx	52.1	39.2	Control 3 Parietal Ctx	31.4	29.3
Control 3 Temporal Ctx	34.9	28.1	Control (Path) 1	48.6	58.6

			Parietal Ctx		
Control 4 Temporal Ctx	62.9	36.3	Control (Path) 2 Parietal Ctx	46.3	27.0
Control (Path) 1 Temporal Ctx	75.8	50.0	Control (Path) 3 Parietal Ctx	26.1	23.8
Control (Path) 2 Temporal Ctx	56.6	41.8	Control (Path) 4 Parietal Ctx	48.0	54.3

Table AFD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3218, Run 168013878	Rel. Exp.(%) Ag3378, Run 165674263	Tissue Name	Rel. Exp.(%) Ag3218, Run 168013878	Rel. Exp.(%) Ag3378, Run 165674263
Liver adenocarcinoma	10.7	20.2	Kidney (fetal)	48.3	13.9
Pancreas	10.8	13.1	Renal ca. 786-0	15.6	10.4
Pancreatic ca. CAPAN 2	9.6	5.4	Renal ca. A498	19.2	14.9
Adrenal gland	5.1	18.4	Renal ca. RXF 393	39.0	33.2
Thyroid	12.3	33.2	Renal ca. ACHN	12.1	11.3
Salivary gland	5.1	5.5	Renal ca. UO-31	18.9	17.8
Pituitary gland	21.5	74.7	Renal ca. TK-10	20.0	10.1
Brain (fetal)	19.5	36.1	Liver	18.0	8.7
Brain (whole)	22.1	29.9	Liver (fetal)	5.5	25.3
Brain (amygdala)	57.4	46.7	Liver ca. (hepatoblast) HepG2	14.2	14.1
Brain (cerebellum)	25.2	23.5	Lung	14.1	18.7
Brain (hippocampus)	28.1	85.9	Lung (fetal)	17.2	4.0
Brain (substantia nigra)	11.5	16.7	Lung ca. (small cell) LX-1	6.5	14.8
Brain (thalamus)	57.0	67.4	Lung ca. (small cell) NCI-H69	20.6	4.8

671

Cerebral Cortex	75.8	36.9	Lung ca. (s.cell var.) SHP-77	**100.0**	39.8
Spinal cord	9.7	13.2	Lung ca. (large cell)NCI-H460	5.0	37.1
glio/astro U87-MG	22.8	13.6	Lung ca. (non-sm. cell) A549	27.7	13.6
glio/astro U-118-MG	37.4	79.6	Lung ca. (non-s.cell) NCI-H23	61.1	44.1
astrocytoma SW1783	29.9	14.9	Lung ca. (non-s.cell) HOP-62	29.9	13.7
neuro*; met SK-N-AS	17.1	52.5	Lung ca. (non-s.cl) NCI-H522	11.3	3.1
astrocytoma SF-539	15.5	16.0	Lung ca. (squam.) SW 900	23.2	13.5
astrocytoma SNB-75	43.8	50.0	Lung ca. (squam.) NCI-H596	41.5	10.2
glioma SNB-19	17.9	26.2	Mammary gland	14.8	35.1
glioma U251	47.6	39.0	Breast ca.* (pl.ef) MCF-7	48.6	39.0
glioma SF-295	12.3	10.7	Breast ca.* (pl.ef) MDA-MB-231	25.9	60.7
Heart (fetal)	38.4	8.0	Breast ca.* (pl.ef) T47D	77.4	21.2
Heart.	3.5	5.0	Breast ca. BT-549	47.0	95.9
Skeletal muscle (fetal)	17.0	10.0	Breast ca. MDA-N	16.6	7.3
Skeletal muscle	4.4	7.2	Ovary	10.1	4.7
Bone marrow	1.3	14.7	Ovarian ca. OVCAR-3	36.3	31.2
Thymus	13.9	12.3	Ovarian ca. OVCAR-4	33.0	20.7
Spleen	2.6	12.9	Ovarian ca. OVCAR-5	42.6	15.7
Lymph node	1.7	15.9	Ovarian ca. OVCAR-8	8.7	5.2

672

Colorectal	18.2	11.8	Ovarian ca. IGROV-1	11.3	15.1
Stomach	14.8	33.7	Ovarian ca.* (ascites) SK-OV-3	43.5	17.0
Small intestine	18.3	66.0	Uterus	10.5	21.8
Colon ca. SW480	12.9	14.2	Placenta	2.6	15.0
Colon ca.* SW620(SW480 met)	17.0	14.2	Prostate	11.7	30.6
Colon ca. HT29	17.2	18.8	Prostate ca.* (bone met)PC-3	35.4	40.3
Colon ca. HCT-116	16.5	18.2	Testis	23.3	**100.0**
Colon ca. CaCo-2	20.9	7.4	Melanoma Hs688(A).T	5.0	1.4
Colon ca. tissue(ODO3866)	14.7	21.9	Melanoma* (met) Hs688(B).T	6.0	3.5
Colon ca. HCC-2998	22.1	13.1	Melanoma UACC-62	14.3	12.2
Gastric ca.* (liver met) NCI-N87	48.6	82.4	Melanoma M14	3.1	8.2
Bladder	6.2	4.7	Melanoma LOX IMVI	30.1	8.4
Trachea	12.8	49.3	Melanoma* (met) SK-MEL-5	21.8	13.1
Kidney	43.5	35.4	Adipose	9.2	3.0

Table AFE. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3218, Run 174416494	Tissue Name	Rel. Exp.(%) Ag3218, Run 174416494
Normal Colon	5.9	Kidney Margin (OD04348)	70.2
Colon cancer (OD06064)	5.6	Kidney malignant cancer (OD06204B)	3.9
Colon Margin (OD06064)	3.6	Kidney normal adjacent tissue (OD06204E)	6.7
Colon cancer (OD06159)	6.3	Kidney Cancer (OD04450-01)	15.1
Colon Margin	7.0	Kidney Margin	3.1

673

(OD06159)		(OD04450-03)	
Colon cancer (OD06297-04)	2.6	Kidney Cancer 8120613	2.5
Colon Margin (OD06297-05)	5.6	Kidney Margin 8120614	18.2
CC Gr.2 ascend colon (ODO3921)	20.0	Kidney Cancer 9010320	2.4
CC Margin (ODO3921)	13.7	Kidney Margin 9010321	4.4
Colon cancer metastasis (OD06104)	0.0	Kidney Cancer 8120607	23.0
Lung Margin (OD06104)	11.0	Kidney Margin 8120608	15.1
Colon mets to lung (OD04451-01)	29.9	Normal Uterus	2.3
Lung Margin (OD04451-02)	0.3	Uterine Cancer 064011	6.1
Normal Prostate	5.6	Normal Thyroid	6.6
Prostate Cancer (OD04410)	3.9	Thyroid Cancer 064010	6.8
Prostate Margin (OD04410)	6.1	Thyroid Cancer A302152	11.9
Normal Ovary	7.0	Thyroid Margin A302153	7.7
Ovarian cancer (OD06283-03)	1.3	Normal Breast	3.4
Ovarian Margin (OD06283-07)	0.0	Breast Cancer (OD04566)	9.9
Ovarian Cancer 064008	31.2	Breast Cancer 1024	16.8
Ovarian cancer (OD06145)	3.5	Breast Cancer (OD04590-01)	**100.0**
Ovarian Margin (OD06145)	8.4	Breast Cancer Mets (OD04590-03)	26.2
Ovarian cancer (OD06455-03)	13.7	Breast Cancer Metastasis (OD04655-05)	36.3
Ovarian Margin (OD06455-07)	1.1	Breast Cancer 064006	5.4
Normal Lung	5.4	Breast Cancer 9100266	12.8
Invasive poor diff. lung adeno (ODO4945-01)	14.5	Breast Margin 9100265	1.0
Lung Margin (ODO4945-03)	2.7	Breast Cancer A209073	3.3
Lung Malignant Cancer (OD03126)	1.8	Breast Margin A2090734	11.7

Tissue Name		Tissue Name	
Lung Margin (OD03126)	5.1	Breast cancer (OD06083)	6.9
Lung Cancer (OD05014A)	12.8	Breast cancer node metastasis (OD06083)	10.7
Lung Margin (OD05014B)	3.3	Normal Liver	9.4
Lung cancer (OD06081)	6.3	Liver Cancer 1026	2.6
Lung Margin (OD06081)	2.7	Liver Cancer 1025	9.7
Lung Cancer (OD04237-01)	12.9	Liver Cancer 6004-T	10.4
Lung Margin (OD04237-02)	6.4	Liver Tissue 6004-N	5.3
Ocular Melanoma Metastasis	4.6	Liver Cancer 6005-T	4.2
Ocular Melanoma Margin (Liver)	0.1	Liver Tissue 6005-N	11.5
Melanoma Metastasis	1.6	Liver Cancer 064003	22.5
Melanoma Margin (Lung)	4.6	Normal Bladder	6.1
Normal Kidney	10.4	Bladder Cancer 1023	10.8
Kidney Ca, Nuclear grade 2 (OD04338)	14.6	Bladder Cancer A302173	15.1
Kidney Margin (OD04338)	10.5	Normal Stomach	15.0
Kidney Ca Nuclear grade 1/2 (OD04339)	44.8	Gastric Cancer 9060397	7.1
Kidney Margin (OD04339)	17.7	Stomach Margin 9060396	10.4
Kidney Ca, Clear cell type (OD04340)	5.3	Gastric Cancer 9060395	8.4
Kidney Margin (OD04340)	25.3	Stomach Margin 9060394	10.4
Kidney Ca, Nuclear grade 3 (OD04348)	7.5	Gastric Cancer 064005	7.7

Table AFF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3218, Run 164682519	Rel. Exp.(%) Ag3378, Run 165296553	Tissue Name	Rel. Exp.(%) Ag3218, Run 164682519	Rel. Exp.(%) Ag3378, Run 165296553
Secondary Th1 act	18.2	25.7	HUVEC IL-1beta	14.5	12.9

675

Secondary Th2 act	39.0	26.6	HUVEC IFN gamma	47.0	25.3
Secondary Tr1 act	33.2	19.1	HUVEC TNF alpha + IFN gamma	43.5	45.1
Secondary Th1 rest	9.5	12.2	HUVEC TNF alpha + IL4	37.1	48.0
Secondary Th2 rest	11.2	5.1	HUVEC IL-11	43.5	18.0
Secondary Tr1 rest	22.7	8.0	Lung Microvascular EC none	16.8	61.6
Primary Th1 act	43.2	27.9	Lung Microvascular EC TNFalpha + IL-1beta	18.6	14.7
Primary Th2 act	30.1	12.0	Microvascular Dermal EC none	31.2	23.8
Primary Tr1 act	24.7	14.4	Microsvasular Dermal EC TNFalpha + IL-1beta	66.4	22.1
Primary Th1 rest	25.2	17.4	Bronchial epithelium TNFalpha + IL1beta	30.6	29.3
Primary Th2 rest	15.5	7.5	Small airway epithelium none	36.1	24.3
Primary Tr1 rest	21.3	6.7	Small airway epithelium TNFalpha + IL-1beta	76.3	62.9
CD45RA CD4 lymphocyte act	35.4	16.6	Coronery artery SMC rest	49.7	28.1
CD45RO CD4 lymphocyte act	27.9	25.9	Coronery artery SMC TNFalpha + IL-1beta	25.3	23.7
CD8 lymphocyte act	21.0	14.8	Astrocytes rest	22.2	31.2
Secondary CD8 lymphocyte rest	39.2	17.8	Astrocytes TNFalpha + IL-1beta	26.1	25.0
Secondary CD8 lymphocyte act	20.9	7.4	KU-812 (Basophil) rest	90.8	85.3
CD4 lymphocyte none	4.5	11.8	KU-812 (Basophil) PMA/ionomycin	87.1	72.2
2ry	2.6	10.0	CCD1106	36.6	36.9

Th1/Th2/Tr1_anti-CD95 CH11			(Keratinocytes) none		
LAK cells rest	11.7	12.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	33.4	20.4
LAK cells IL-2	6.8	27.5	Liver cirrhosis	25.5	19.9
LAK cells IL-2+IL-12	37.1	11.6	Lupus kidney	44.4	15.7
LAK cells IL-2+IFN gamma	20.7	19.1	NCI-H292 none	79.6	64.6
LAK cells IL-2+ IL-18	21.9	14.7	NCI-H292 IL-4	85.3	96.6
LAK cells PMA/ionomycin	4.7	3.3	NCI-H292 IL-9	**100.0**	98.6
NK Cells IL-2 rest	11.9	11.3	NCI-H292 IL-13	68.8	50.7
Two Way MLR 3 day	23.7	11.0	NCI-H292 IFN gamma	80.1	56.6
Two Way MLR 5 day	12.5	6.1	HPAEC none	38.4	27.2
Two Way MLR 7 day	12.3	8.7	HPAEC TNF alpha + IL-1 beta	42.6	43.2
PBMC rest	6.0	5.7	Lung fibroblast none	31.2	21.3
PBMC PWM	40.3	27.7	Lung fibroblast TNF alpha + IL-1 beta	14.7	24.5
PBMC PHA-L	37.9	17.7	Lung fibroblast IL-4	47.0	42.6
Ramos (B cell) none	11.7	14.9	Lung fibroblast IL-9	49.3	30.6
Ramos (B cell) ionomycin	33.9	26.8	Lung fibroblast IL-13	36.6	42.6
B lymphocytes PWM	33.7	40.9	Lung fibroblast IFN gamma	44.8	22.5
B lymphocytes CD40L and IL-4	34.4	18.3	Dermal fibroblast CCD1070 rest	33.7	47.3
EOL-1 dbcAMP	50.0	28.1	Dermal fibroblast CCD1070 TNF alpha	47.3	33.2
EOL-1 dbcAMP PMA/ionomycin	44.1	32.1	Dermal fibroblast CCD1070 IL-1 beta	50.0	34.6
Dendritic cells none	33.9	19.6	Dermal fibroblast IFN gamma	24.0	34.4
Dendritic cells LPS	21.9	10.2	Dermal fibroblast	24.3	32.8

677

			IL-4		
Dendritic cells anti-CD40	49.7	33.9	IBD Colitis 2	6.0	11.7
Monocytes rest	10.7	10.3	IBD Crohn's	25.3	26.1
Monocytes LPS	30.6	9.3	Colon	70.7	**100.0**
Macrophages rest	41.2	33.7	Lung	64.6	17.7
Macrophages LPS	20.0	7.5	Thymus	80.7	56.3
HUVEC none	26.8	29.5	Kidney	26.4	41.5
HUVEC starved	26.2	37.6			

CNS_neurodegeneration_v1.0 Summary: Ag3218/Ag3378 - Two different experiments using probe/primer sets with the same sequence are in very good agreement. This panel confirms the expression of this gene at low levels to moderate levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.3D for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

Panel 1.3D Summary: Ag3218/Ag3378 - Two different experiments using probe/primer sets with the same sequence are in good agreement. Highest expression is seen in testis and a lung cancer cell line (CTs=30-31). This panel confirms the expression of this gene at low levels in the brain. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

This gene product is also expressed in adipose, pancreas, thyroid, pituitary, heart, and liver. This widespread expression in tissues with metabolic function suggests that this gene product may be important for the pathogenesis, diagnosis, and/or treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

Based on expression in this panel, this gene may be involved in gastric, pancreatic, brain, colon, renal, lung, breast, ovarian and prostate cancer as well as melanomas. Thus, expression of this gene could be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition using antibodies or small molecule drugs might be of use in the treatment of these cancers.

Panel 2.2 Summary: Ag3218 - This gene is expressed at low to moderate levels in many samples on this panel, with the highest levels of expression in breast cancer sample OD04590-01 (CT=30.3). This gene is expressed in a cluster of breast cancer samples with no expression in normal breast (CT>35). Similarly, this gene is expressed in ovarian cancer samples at higher levels than the matched margin samples.

Interestingly, this gene is expressed at higher levels in kidney cancer margin samples than in the matched cancer samples.

This gene is homologous to a mouse myosin phosphatase targeting subunit (MYPT) which have been found to play a role in cell division. MYPT undergoes mitosis-specific phosphorylation which is reversed during cytokinesis.

References:

1. Totsukawa G, Yamakita Y, Yamashiro S, Hosoya H, Hartshorne DJ, Matsumura F. Activation of myosin phosphatase targeting subunit by mitosis-specific phosphorylation. J Cell Biol 1999 Feb 22;144(4):735-44.

Panel 4D Summary: Ag3218/Ag3378 - Two different experiments using probe/primer sets with the same sequence are in very good agreement. Highest expression is seen in the colon and a mucoepidermoid cell line (CTs=30-32). This gene is expressed at low to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

AG. CG59241-01: Amiloride-sensitive sodium channel

Expression of gene CG59241-01 was assessed using the primer-probe set Ag3407, described in Table AGA. Results of the RTQ-PCR runs are shown in Tables AGB, AGC and AGD.

Table AGA. Probe Name Ag3407

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtcaccctctgcaacactaatg-3'	22	268	496
Probe	TET-5'-ctgtcccagctcagctaccctgactt-3'-TAMRA	26	298	497
Reverse	5'-tttcatccagtcccagcat-3'	19	340	498

5   Table AGB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3407, Run 210349883	Tissue Name	Rel. Exp.(%) Ag3407, Run 210349883
AD 1 Hippo	18.4	Control (Path) 3 Temporal Ctx	4.1
AD 2 Hippo	29.7	Control (Path) 4 Temporal Ctx	40.3
AD 3 Hippo	18.3	AD 1 Occipital Ctx	36.9
AD 4 Hippo	5.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	91.4	AD 3 Occipital Ctx	19.1
AD 6 Hippo	80.7	AD 4 Occipital Ctx	18.8
Control 2 Hippo	9.3	AD 5 Occipital Ctx	18.3
Control 4 Hippo	19.9	AD 6 Occipital Ctx	28.9
Control (Path) 3 Hippo	8.8	Control 1 Occipital Ctx	4.3
AD 1 Temporal Ctx	28.5	Control 2 Occipital Ctx	80.1
AD 2 Temporal Ctx	41.8	Control 3 Occipital Ctx	20.2
AD 3 Temporal Ctx	32.5	Control 4 Occipital Ctx	6.0
AD 4 Temporal Ctx	36.3	Control (Path) 1 Occipital Ctx	92.7
AD 5 Inf Temporal	**100.0**	Control (Path) 2	25.3

Ctx		Occipital Ctx	
AD 5 SupTemporal Ctx	56.6	Control (Path) 3 Occipital Ctx	3.0
AD 6 Inf Temporal Ctx	82.4	Control (Path) 4 Occipital Ctx	41.2
AD 6 Sup Temporal Ctx	44.1	Control 1 Parietal Ctx	21.9
Control 1 Temporal Ctx	15.3	Control 2 Parietal Ctx	79.0
Control 2 Temporal Ctx	24.1	Control 3 Parietal Ctx	22.2
Control 3 Temporal Ctx	34.6	Control (Path) 1 Parietal Ctx	77.9
Control 4 Temporal Ctx	12.0	Control (Path) 2 Parietal Ctx	47.6
Control (Path) 1 Temporal Ctx	53.6	Control (Path) 3 Parietal Ctx	6.2
Control (Path) 2 Temporal Ctx	56.6	Control (Path) 4 Parietal Ctx	67.4

Table AGC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3407, Run 216821458	Tissue Name	Rel. Exp.(%) Ag3407, Run 216821458
Adipose	0.2	Renal ca. TK-10	16.6
Melanoma* Hs688(A).T	2.3	Bladder	0.3
Melanoma* Hs688(B).T	0.4	Gastric ca. (liver met.) NCI-N87	8.8
Melanoma* M14	2.0	Gastric ca. KATO III	0.7
Melanoma* LOXIMVI	2.5	Colon ca. SW-948	3.7
Melanoma* SK-MEL-5	8.7	Colon ca. SW480	14.1
Squamous cell carcinoma SCC-4	1.2	Colon ca.* (SW480 met) SW620	21.2
Testis Pool	0.4	Colon ca. HT29	10.7
Prostate ca.* (bone met) PC-3	4.4	Colon ca. HCT-116	64.2
Prostate Pool	2.3	Colon ca. CaCo-2	32.3
Placenta	0.5	Colon cancer tissue	13.2
Uterus Pool	0.0	Colon ca. SW1116	12.5
Ovarian ca. OVCAR-3	8.4	Colon ca. Colo-205	0.3

Ovarian ca. SK-OV-3	9.7	Colon ca. SW-48	0.6
Ovarian ca. OVCAR-4	1.6	Colon Pool	2.8
Ovarian ca. OVCAR-5	18.9	Small Intestine Pool	4.5
Ovarian ca. IGROV-1	4.9	Stomach Pool	1.4
Ovarian ca. OVCAR-8	5.9	Bone Marrow Pool	1.8
Ovary	2.0	Fetal Heart	2.4
Breast ca. MCF-7	16.7	Heart Pool	0.3
Breast ca. MDA-MB-231	12.1	Lymph Node Pool	3.5
Breast ca. BT 549	22.7	Fetal Skeletal Muscle	1.9
Breast ca. T47D	27.4	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	4.5	Spleen Pool	0.0
Breast Pool	2.9	Thymus Pool	2.1
Trachea	9.0	CNS cancer (glio/astro) U87-MG	0.9
Lung	0.0	CNS cancer (glio/astro) U-118-MG	11.7
Fetal Lung	10.8	CNS cancer (neuro;met) SK-N-AS	58.6
Lung ca. NCI-N417	1.3	CNS cancer (astro) SF-539	28.1
Lung ca. LX-1	21.8	CNS cancer (astro) SNB-75	24.7
Lung ca. NCI-H146	5.4	CNS cancer (glio) SNB-19	7.3
Lung ca. SHP-77	11.7	CNS cancer (glio) SF-295	4.8
Lung ca. A549	8.0	Brain (Amygdala) Pool	3.9
Lung ca. NCI-H526	0.0	Brain (cerebellum)	36.1
Lung ca. NCI-H23	7.4	Brain (fetal)	**100.0**
Lung ca. NCI-H460	5.4	Brain (Hippocampus) Pool	5.6
Lung ca. HOP-62	2.9	Cerebral Cortex Pool	5.6
Lung ca. NCI-H522	8.5	Brain (Substantia nigra) Pool	7.1
Liver	0.0	Brain (Thalamus) Pool	11.3
Fetal Liver	0.0	Brain (whole)	13.4
Liver ca. HepG2	0.8	Spinal Cord Pool	12.7
Kidney Pool	2.1	Adrenal Gland	0.0

682

Fetal Kidney	3.7	Pituitary gland Pool	0.0
Renal ca. 786-0	1.7	Salivary Gland	0.9
Renal ca. A498	0.7	Thyroid (female)	0.0
Renal ca. ACHN	1.9	Pancreatic ca. CAPAN2	2.3
Renal ca. UO-31	0.2	Pancreas Pool	2.6

Table AGD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3407, Run 165296462	Tissue Name	Rel. Exp.(%) Ag3407, Run 165296462
Secondary Th1 act	7.9	HUVEC IL-1beta	0.0
Secondary Th2 act	17.1	HUVEC IFN gamma	0.0
Secondary Tr1 act	40.1	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	4.4	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	7.0	HUVEC IL-11	0.0
Secondary Tr1 rest	11.7	Lung Microvascular EC none	0.0
Primary Th1 act	61.1	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	69.3	Microvascular Dermal EC none	0.0
Primary Tr1 act	90.8	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	20.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	42.6	Small airway epithelium none	3.0
Primary Tr1 rest	52.5	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	2.8	Coronery artery SMC rest	3.6
CD45RO CD4 lymphocyte act	14.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	5.8	Astrocytes rest	11.6
Secondary CD8 lymphocyte rest	18.9	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	22.2	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-	4.5	CCD1106	2.7

CD95 CH11		(Keratinocytes) none	
LAK cells rest	3.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	4.0	Liver cirrhosis	13.5
LAK cells IL-2+IL-12	5.7	Lupus kidney	4.1
LAK cells IL-2+IFN gamma	21.3	NCI-H292 none	9.0
LAK cells IL-2+ IL-18	6.7	NCI-H292 IL-4	14.8
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	3.5
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	5.0	NCI-H292 IFN gamma	5.5
Two Way MLR 5 day	2.3	HPAEC none	0.0
Two Way MLR 7 day	8.2	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	2.8
PBMC PWM	21.3	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	20.4	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	2.8	Lung fibroblast IL-13	0.0
B lymphocytes PWM	**100.0**	Lung fibroblast IFN gamma	1.4
B lymphocytes CD40L and IL-4	19.8	Dermal fibroblast CCD1070 rest	34.4
EOL-1 dbcAMP	2.6	Dermal fibroblast CCD1070 TNF alpha	68.8
EOL-1 dbcAMP PMA/ionomycin	6.2	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	14.1
Dendritic cells anti-CD40	6.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	6.5	Colon	42.3
Macrophages rest	0.0	Lung	35.8
Macrophages LPS	0.0	Thymus	45.4
HUVEC none	0.0	Kidney	55.1
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag3407 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.4 Summary: Ag3407 Highest expression of the CG59241-01 gene is seen in fetal brain (CT=31.3). Furthermore, low to moderate levels of expression is also observed in CNS cancer cell lines (CTs=32-34). The CG59241-01 gene codes for a putative amiloride-sensitive sodium channel. A similar amiloride-sensitive sodium channel was shown to be highly expressed in malignant glioblastoma multiforme tumors and to be a charachteristic feature of malignant brain tumor cells (Ref.1). Therefore, therapeutic modulation of the activity of the protein encoded by this gene may be beneficial in the treatment of CNS cancer. Significant expression is also seen in a cluster of cell lines derived from brain, colon, breast, and ovarian cancers. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of these cancers.

In addition, this gene is expressed at low levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

References:

1. Bubien JK, Keeton DA, Fuller CM, Gillespie GY, Reddy AT, Mapstone TB, Benos DJ. (1999) Malignant human gliomas express an amiloride-sensitive Na+ conductance. Am J Physiol 276(6 Pt 1):C1405-10

Panel 4D Summary: Ag3407 Highest expression Of the CG59241-01 gene is detected in PWM treated B lymphocytes (CT=32). Similar expression is also detected in primary activated Th1, Th2 and Tr1 cells, as well as TNF alpha treated dermal fibroblast CCD1070 cells (CTs=32). Therefore, expression of this gene can be used to distinguish these samples from other samples in the panel. Furthermore, this gene is expressed in activated

lymphocytes. Likewise, no expression of this gene is seen in PBMC that contain normal B cells (CT=40), but it is induced when PBMC are treated with the pokeweed mitogen or PHA-L (CTs=34). In addition, the transcript is not seen in the B cell lymphoma Ramos regardless of stimulation. Therefore, the gene product could potentially be used

5   therapeutically in the treatment of Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, psoriasis and in other diseases in which T cells and B cells are activated.

In addition, low expression of this gene is also observed in normal colon, lung, thymus and kidney tissues. The CG59241-01 gene encodes an amiloride-sensitive sodium

10   channel. A similar channel, the amiloride-sensitive epithelial sodium channel (ENaC) constitutes the limiting step for sodium reabsorption in epithelial cells that line the distal nephron, distal colon, ducts of several exocrine glands and lung airways and plays an important role in pathophysiological and clinical conditions such as hypertension or lung edema. ENaC has been implicated in two genetic diseases, Liddle's syndrome and

15   pseudohypoaldosteronism (PHA-1) (Ref.1). Therefore, antibody or small molecule therapies designed with the protein encoded for by CG59241-01 gene could modulate kidney/colon/lung function and be important in the treatment of inflammatory or autoimmune diseases of these tissues in addition to hypertension, lung edema, Liddle's syndrom and PHA-1.

20   **Reference.**

1. Hummler E. (1998) Reversal of convention: from man to experimental animal in elucidating the function of the renal amiloride-sensitive sodium channel. Exp Nephrol 1998 Jul-Aug;6(4):265-71

AH. CG58602-01: FAD binding domain containing protein

25   Expression of gene CG58602-01 was assessed using the primer-probe set Ag3385, described in Table AHA. Results of the RTQ-PCR runs are shown in Tables AHB, AHC and AHD.

Table AHA. Probe Name Ag3385

Primers	Sequences	Length	Start Position	SEQ ID NO:

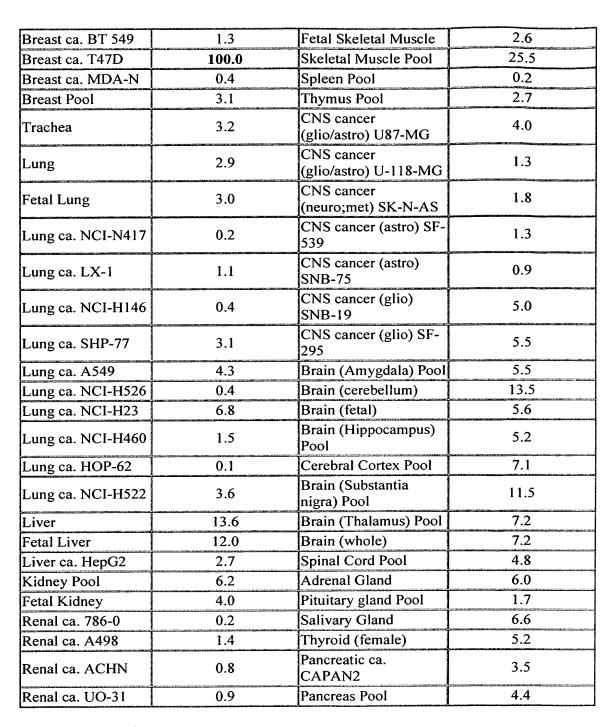| Forward	5'-tcatgaatccaggcaaagtg-3'	20	1427	499
Probe	TET-5'-ttagcccacaagttccctgactacgg-3'-TAMRA	26	1468	500
Reverse	5'-tggcatgaagaaaagttcca-3'	20	1503	501

Table AHB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3385, Run 210154892	Tissue Name	Rel. Exp.(%) Ag3385, Run 210154892
AD 1 Hippo	34.6	Control (Path) 3 Temporal Ctx	21.2
AD 2 Hippo	47.6	Control (Path) 4 Temporal Ctx	36.1
AD 3 Hippo	11.9	AD 1 Occipital Ctx	28.1
AD 4 Hippo	24.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	56.3	AD 3 Occipital Ctx	15.0
AD 6 Hippo	63.3	AD 4 Occipital Ctx	34.9
Control 2 Hippo	42.6	AD 5 Occipital Ctx	52.1
Control 4 Hippo	24.7	AD 6 Occipital Ctx	25.3
Control (Path) 3 Hippo	23.3	Control 1 Occipital Ctx	14.3
AD 1 Temporal Ctx	23.8	Control 2 Occipital Ctx	69.3
AD 2 Temporal Ctx	73.7	Control 3 Occipital Ctx	29.5
AD 3 Temporal Ctx	7.3	Control 4 Occipital Ctx	14.9
AD 4 Temporal Ctx	39.0	Control (Path) 1 Occipital Ctx	68.3
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	11.0
AD 5 Sup Temporal Ctx	55.5	Control (Path) 3 Occipital Ctx	8.9
AD 6 Inf Temporal Ctx	64.2	Control (Path) 4 Occipital Ctx	17.3
AD 6 Sup Temporal Ctx	54.0	Control 1 Parietal Ctx	32.8
Control 1 Temporal Ctx	23.8	Control 2 Parietal Ctx	62.0
Control 2 Temporal Ctx	50.3	Control 3 Parietal Ctx	33.4
Control 3 Temporal Ctx	38.4	Control (Path) 1 Parietal Ctx	70.7
Control 3	19.2	Control (Path) 2	31.4

Temporal Ctx		Parietal Ctx	
Control (Path) 1 Temporal Ctx	56.6	Control (Path) 3 Parietal Ctx	20.9
Control (Path) 2 Temporal Ctx	47.6	Control (Path) 4 Parietal Ctx	43.2

Table AHC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3385, Run 217043538	Tissue Name	Rel. Exp.(%) Ag3385, Run 217043538
Adipose	2.4	Renal ca. TK-10	3.5
Melanoma* Hs688(A).T	0.7	Bladder	6.6
Melanoma* Hs688(B).T	1.1	Gastric ca. (liver met.) NCI-N87	2.1
Melanoma* M14	0.9	Gastric ca. KATO III	0.9
Melanoma* LOXIMVI	1.3	Colon ca. SW-948	4.5
Melanoma* SK-MEL-5	2.2	Colon ca. SW480	0.8
Squamous cell carcinoma SCC-4	0.1	Colon ca.* (SW480 met) SW620	1.3
Testis Pool	1.3	Colon ca. HT29	0.6
Prostate ca.* (bone met) PC-3	5.8	Colon ca. HCT-116	1.9
Prostate Pool	4.0	Colon ca. CaCo-2	28.5
Placenta	2.5	Colon cancer tissue	2.0
Uterus Pool	0.5	Colon ca. SW1116	0.9
Ovarian ca. OVCAR-3	1.1	Colon ca. Colo-205	3.5
Ovarian ca. SK-OV-3	3.7	Colon ca. SW-48	4.2
Ovarian ca. OVCAR-4	0.2	Colon Pool	3.0
Ovarian ca. OVCAR-5	42.0	Small Intestine Pool	3.5
Ovarian ca. IGROV-1	8.0	Stomach Pool	1.8
Ovarian ca. OVCAR-8	2.7	Bone Marrow Pool	0.9
Ovary	3.3	Fetal Heart	12.9
Breast ca. MCF-7	10.3	Heart Pool	8.3
Breast ca. MDA-MB-231	3.0	Lymph Node Pool	3.5

Breast ca. BT 549	1.3	Fetal Skeletal Muscle	2.6
Breast ca. T47D	**100.0**	Skeletal Muscle Pool	25.5
Breast ca. MDA-N	0.4	Spleen Pool	0.2
Breast Pool	3.1	Thymus Pool	2.7
Trachea	3.2	CNS cancer (glio/astro) U87-MG	4.0
Lung	2.9	CNS cancer (glio/astro) U-118-MG	1.3
Fetal Lung	3.0	CNS cancer (neuro;met) SK-N-AS	1.8
Lung ca. NCI-N417	0.2	CNS cancer (astro) SF-539	1.3
Lung ca. LX-1	1.1	CNS cancer (astro) SNB-75	0.9
Lung ca. NCI-H146	0.4	CNS cancer (glio) SNB-19	5.0
Lung ca. SHP-77	3.1	CNS cancer (glio) SF-295	5.5
Lung ca. A549	4.3	Brain (Amygdala) Pool	5.5
Lung ca. NCI-H526	0.4	Brain (cerebellum)	13.5
Lung ca. NCI-H23	6.8	Brain (fetal)	5.6
Lung ca. NCI-H460	1.5	Brain (Hippocampus) Pool	5.2
Lung ca. HOP-62	0.1	Cerebral Cortex Pool	7.1
Lung ca. NCI-H522	3.6	Brain (Substantia nigra) Pool	11.5
Liver	13.6	Brain (Thalamus) Pool	7.2
Fetal Liver	12.0	Brain (whole)	7.2
Liver ca. HepG2	2.7	Spinal Cord Pool	4.8
Kidney Pool	6.2	Adrenal Gland	6.0
Fetal Kidney	4.0	Pituitary gland Pool	1.7
Renal ca. 786-0	0.2	Salivary Gland	6.6
Renal ca. A498	1.4	Thyroid (female)	5.2
Renal ca. ACHN	0.8	Pancreatic ca. CAPAN2	3.5
Renal ca. UO-31	0.9	Pancreas Pool	4.4

Table AHD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3385, Run 165296471	Tissue Name	Rel. Exp.(%) Ag3385, Run 165296471
Secondary Th1 act	1.2	HUVEC IL-1beta	0.0
Secondary Th2 act	3.6	HUVEC IFN gamma	3.7